AN INTRODUCTION TO GENERALIZED LINEAR MIXED MODELS

Stephen D. Kachman
Department of Biometry, University of Nebraska–Lincoln

Abstract

Linear mixed models provide a powerful means of predicting breeding values. However, for many traits of economic importance the assumptions of linear responses, constant variance, and normality are questionable. Generalized linear mixed models provide a means of modeling these deviations from the usual linear mixed model. This paper will examine what constitutes a generalized linear mixed model, issues involved in constructing a generalized linear mixed model, and the modifications necessary to convert a linear mixed model program into a generalized linear mixed model program.

1 Introduction

Generalized linear mixed models (GLMM) [1, 2, 3, 6] have attracted considerable attention over the years. With the advent of SAS’s GLIMMIX macro [5], generalized linear mixed models have become available to a larger audience. However, in a typical breeding evaluation generic packages are too inefficient and implementations in FORTRAN or C are needed. In addition, GLMM’s pose additional challenges with some solutions heading for \( \pm \infty \).

The objective of this paper is to provide an introduction to generalized linear mixed models. In section 2, I will discuss some of the deficiencies of a linear model. In section 3, I will present the generalized linear mixed model. In section 4, I will present the estimation equations for the fixed and random effects. In section 5, I will present a set of estimating equations for the variance components. In section 6, I will discuss some of the computational issues involved when these approaches are used in practice.

2 Mixed models

In this section I will discuss the linear mixed model and when the implied assumptions are not appropriate. A linear mixed model is

\[
y|u \sim N(X\beta + Zu, R)
\]

where \( u \sim N(0, G) \), \( X \) and \( Z \) are known design matrices, and the covariance matrices \( R \) and \( G \) may depend on a set of unknown variance components. The linear mixed model assumes that the relationship between the mean of the dependent variable \( y \) and the fixed and random effects can be modeled as a linear function, the variance is not a
function of the mean, and that the random effects follow a normal distribution. Any or all these assumptions may be violated for certain traits.

A case where the assumption of linear relationships is questionable is pregnancy rate. Pregnancy is a zero/one trait, that is, at a given point an animal is either pregnant (1) or is not pregnant (0). For example, a change in management that is expected to increase pregnancy rate by .1 in a herd with a current pregnancy rate of .5 would be expected to have a smaller effect in a herd with a current pregnancy rate of .8; that is, a treatment, an environmental factor, or a sire would be expected to have a larger effect when the mean pregnancy rate is .5 than when the pregnancy rate is .8.

Another case where the assumption of a linear relationship is questionable is the analysis of growth. Cattle, pigs, sheep, and mice have similar growth curves over time; that is, after a period of rapid growth they reach maturity and the growth rate is considerably slower. The relationship of time with weight is not linear, with time having a much larger effect when the animal is young and a very small effect when the animal is mature.

The assumption of constant variance is also questionable for pregnancy rate. When the predicted pregnancy rate, $\mu$, for a cow is .5 the variance is $\mu (1 - \mu) = .25$. If on the other hand the predicted pregnancy rate for a cow is .8 the variance drops to .16. For some production traits the variance increases as mean level of production increases.

The assumption of normality is also questionable for pregnancy rate. It is difficult to justify the assumption that the density function of a random variable which only takes on two values is similar to a continuous bell shaped curve with values ranging from $-\infty$ to $+\infty$.

A number of approaches have been taken to address the deficiencies of a linear mixed model. Transformations have been used to stabilize the variance, to obtain a linear relationship, and to normalize the distribution. However the transformation needed to stabilize the variance may not be the same transformation needed to obtain a linear relationship. For example a log transformation to stabilize the variance has the side effect that the model on the original scale is multiplicative. Linear and multiplicative adjustments are used to adjust to a common base and to account for heterogeneous variances. Multiple trait analysis can be used to account for heterogeneity of responses in different environments. Separate variances can be estimated for different environmental groups where the environmental groups are based on the observed production. A final option is to ignore the deficiencies of the linear mixed model and proceed as if a linear model does hold.

The above options have the appeal that they are relatively simple and cheap to implement. Given the robustness of the estimation procedures, they can be expected to produce reasonable results. However, these options sidestep the issue that the linear mixed model is incorrect. Specifically we have a set of estimation procedures which are based on a linear mixed model and manipulate the data to make it fit a linear mixed model. It seems more reasonable to start with an appropriate model for the data and use an estimation procedure derived from that model. A generalized linear mixed model is a model which gives us extra flexibility in developing an appropriate model for the data [1].
3 A Generalized Linear Mixed Model

In this section I will present a formulation of a generalized linear mixed model. It differs from presentations such as [1] in that it focuses more on the inverse link function rather than the link function to model the relationship between the linear predictor and the conditional mean. Generalized linear mixed models also includes the nonlinear mixed models of [4].

Figure 1 provides a symbolic representation of a generalized linear mixed model. As in a linear mixed model, a generalized linear mixed model includes fixed effects, $\beta$, with (e.g., management effect); random effects, $u \sim N(0, G)$, (e.g., breeding values); design matrices $X$ and $Z$; and a vector of observations, $y$, (e.g., pregnancy status) for which the conditional distribution given the random effects has mean, $\mu$ (e.g., mean pregnancy rate), and covariance matrix, $R$, (e.g., variance of pregnancy status is $\mu(1-\mu)$).

In addition, a generalized linear mixed model includes a linear predictor, $\eta$, and a link and/or inverse link function. In addition, the conditional mean, $\mu$, depends on the linear predictor through an inverse link function, $h(\cdot)$, and the covariance matrix, $R$, depends on $\mu$ through a variance function.

For example, the mean pregnancy rate, $\mu$, depends on the effect of management and the breeding value of the animal. The management effect and breeding value act
additively on a conceptual underlying scale. Their combined effect on the underlying scale is expressed as the linear predictor, \( \eta \). The linear predictor is then transformed to the observed scale (i.e., mean pregnancy rate) through an inverse link function, \( h(\eta) \). A typical transformation would be

\[
h(\eta) = \frac{e^\eta}{1 + e^\eta}.
\]

### 3.1 Linear Predictor, \( \eta \)

As with a linear mixed model, the fixed and random effects are combined to form a linear predictor

\[
\eta = X\beta + Zu.
\]

With a linear mixed model the model for the vector of observations \( y \) is obtained by adding a vector of residuals, \( e \sim N(0, R) \), as follows

\[
y = \eta + e = X\beta + Zu + e.
\]

Equivalently, the residual variability can be modeled as

\[
y|u \sim N(\eta, R).
\]

Unless \( y \) has a normal distribution, the formulation using \( e \) is clumsy. Therefore, a generalized linear mixed model uses a second approach to model the residual variability. The relationship between the linear predictor and the vector of observations in a generalized linear mixed model is modeled as

\[
y|u \sim (h(\eta), R)
\]

where the notation, \( y|u \sim (h(\eta)R) \), specifies that the conditional distribution of \( y \) given \( u \) has mean, \( h(\eta) \), and variance, \( R \). The conditional distribution of \( y \) given \( u \) will be referred to as the error distribution. Choice of which fixed and random effects to include in the model will follow the same considerations as for a linear mixed model.

It is important to note that the effect of the linear predictor is expressed through an inverse link function. Except for the identity link function, \( h(\eta) = \eta \), the effect of a one unit change in \( \eta \) will not correspond to a one unit change in the conditional mean; that is, predicted progeny difference will depend on the progeny’s environment through \( h(\eta) \). The relationship between the linear predictor and the mean response on the observed scale will be considered in more detail in the next section.

### 3.2 (Inverse) Link Function

The inverse link function is used to map the value of the linear predictor for observation \( i, \eta_i, \) to the conditional mean for observation \( i, \mu_i \). For many traits the inverse link function is one to one, that is both \( \mu_i \) and \( \eta_i \) are scalars. For threshold models, \( \mu_i \) is a \( t \times 1 \) vector,
Table 1: Common link functions and variance functions for various distributions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Link</th>
<th>Inverse Link</th>
<th>$v(\mu)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Identity</td>
<td>$\eta$</td>
<td>1</td>
</tr>
<tr>
<td>Binomial/n</td>
<td>Logit</td>
<td>$e^\eta/(1 + e^\eta)$</td>
<td>$\mu(1 - \mu)/n$</td>
</tr>
<tr>
<td></td>
<td>Probit</td>
<td>$\Phi(\eta)$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>Poisson</td>
<td>Log</td>
<td>$e^\eta$</td>
<td>$\mu$</td>
</tr>
<tr>
<td>Gamma</td>
<td>Inverse</td>
<td>$1/\eta$</td>
<td>$\mu^2$</td>
</tr>
<tr>
<td></td>
<td>Log</td>
<td>$e^\eta$</td>
<td>$\mu$</td>
</tr>
</tbody>
</table>

where $t$ is the number of ordinal levels. For growth curve models, $\mu_i$ is an $n_i \times 1$ vector and $\eta_i$ is a $p \times 1$ vector, where the animal is measured $n_i$ times and there are $p$ growth curve parameters.

For the linear mixed model, the inverse link function is the identity function $h(\eta_i) = \eta_i$. For zero/one traits a logit link function $\eta_i = \ln(\mu_i/(1 - \mu_i))$ is often used, the corresponding inverse link function is $\mu_i = e^{\eta_i}/(1 + e^{\eta_i})$. The logit link function, unlike the identity link function, will always yield estimated means in the range of zero to one. However, the effect of a one unit change in the linear predictor is not constant. When the linear predictor is 0 the corresponding mean is .5. Increasing the linear predictor by 1 to 1 increases the corresponding mean by .23. If the linear predictor is 3, the corresponding mean is .95. Increasing the linear predictor by 1 to 4 increases the corresponding mean by only .03. For most univariate link functions, link and inverse link functions are increasing monotonic functions. In other words, an increase in the linear predictor results in an increase in the conditional mean, but not at a constant rate.

Selection of inverse link functions is typically based on the error distribution. Table 1 lists a number of common distributions along with their link functions. Other considerations include simplicity and the ability to interpret the results of the analysis.

### 3.3 Variance Function

The variance function is used to model non-systematic variability. Typically with a generalized linear model, residual variability arises from two sources. First, variability arises from the sampling distribution. For example, a Poisson random variable with mean $\mu$ has a variance of $\mu$. Second, additional variability, or over-dispersion, is often observed.

Modeling variability due to the sampling distribution is straightforward. In Table 1 the variance functions for some common sampling distributions are given.

Variability due to over-dispersion can be modeled in a number of ways. One approach is to scale the residual variability as $\text{var}(y_i|u) = \phi v(\mu_i)$, where $\phi$ is an over-dispersion parameter. A second approach is to add a additional random effect, $e_i \sim N(0, \phi)$, to the linear predictor for each observation. A third approach is to select another distribution. For example, instead of using a one parameter ($\mu$) Poisson distribution for count data, a two parameter ($\mu, \phi$) negative binomial distribution could be used. The three approaches all involve the estimation of an additional parameter, $\phi$. Scaling the residual variability is the simplest approach, but can yield poor results. The addition of a random effect has the
effect of greatly increasing the computational costs.

3.4 The parts

To summarize, a generalized linear model is composed of three parts. First, a linear predictor, \( \eta = X\beta + Zu \), is used to model the relationship between the fixed and random effects. The residual variability contained in the residual, \( e \), of the linear mixed model equation is incorporated in the variance function of the generalized linear mixed model. Second, an inverse link function, \( \mu_i = h(\eta_i) \), is used to model the relationship between the linear predictor and the conditional mean of the observed trait. In general, the link function is selected to be both simple and reasonable. Third, a variance function, \( v(\mu_i, \phi) \), is used to model the residual variability. Selection of the variance function is typically dictated by the error distribution that was chosen. In addition, observed residual variability is often greater than expected due to sampling and needs to be accounted for with an overdispersion parameter.

4 Estimation and Prediction

The estimating equations for a generalized linear mixed model can be derived in a number of ways. From a Bayesian perspective the solutions to the estimating equations are posterior mode predictors. The estimating equations can be obtained by using a Laplacian approximation of the likelihood. The estimating equations for the fixed and random effects are

\[
\begin{pmatrix}
X'HR^{-1}HX & X'HR^{-1}HZ \\
Z'HR^{-1}HX & Z'HR^{-1}HZ + G^{-1}
\end{pmatrix}
\begin{pmatrix}
\hat{\beta} \\
\hat{u}
\end{pmatrix} =
\begin{pmatrix}
X'H^{-1}y^* \\
Z'H^{-1}y^*
\end{pmatrix}
\tag{1}
\]

where

\[
H = \frac{\partial \mu}{\partial \eta'}
\]

\[
R = \text{var}(y|u)
\]

\[
y^* = y - \mu + H\eta.
\]

The first thing to note is the similarity to the usual mixed model equations. This is easier to see if we rewrite the equations in the following form

\[
\begin{pmatrix}
X'WX & X'WZ \\
Z'WX & Z'WZ + G^{-1}
\end{pmatrix}
\begin{pmatrix}
\hat{\beta} \\
\hat{u}
\end{pmatrix} =
\begin{pmatrix}
X'hry \\
Z'hry
\end{pmatrix}
\tag{2}
\]

where

\[
W = H'R^{-1}H
\]

\[
hry = H'R^{-1}y^*.
\]

Unlike the mixed model equations, the estimating equations (1) for a generalized linear mixed must be solved iteratively.
4.1 Univariate Logit

To see how all the pieces fit together, we will examine a univariate binomial with a logit link function which would be an appropriate model for examining proportion data. Let \( y_i \) be the proportion out of \( n \) for observation \( i \), a reasonable error distribution for \( n y_i \) would be Binomial with parameters, \( n \) and \( \mu_i \). For a univariate model, the \( H, R, \) and \( W \) matrices are all diagonal with diagonal elements equal to \( \frac{\partial \mu_i}{\partial \eta_i} \), \( \nu(\mu_i) \), and \( W_{ii} \) respectively.

The variance function for a scaled Binomial random variable is \( \nu(\mu_i) = \mu_i(1 - \mu_i)/n \). The inverse link function is selected to model the nonlinear response of the means, \( \mu_i \), to changes in the linear predictor, \( \eta_i \). The inverse link function selected is \( \mu_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \). The nonlinear relationship between the linear predictor and the mean can be seen in Figure 2. Changes in the linear predictor when the mean is close to .5 have a larger impact than similar changes when the mean is close to 0 or 1. For example, a change in the linear predictor from 0 to .5 changes the mean from 50% to 62%. However, a change from 2 to 2.5 changes the mean from 88% to 92%.

Figure 2: Inverse logit link function \( \mu = \frac{e^{\eta}}{1 + e^{\eta}} \).
The weight, \( W_{ii} \), and scaled dependent variable, \( h_{ryi} \), for observation \( i \) are

\[
\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial e^{\eta_i}}{\partial \eta_i} = \mu_i(1 - \mu_i)
\]

\[
W_{ii} = \frac{\partial \mu_i}{\partial \eta_i} \left[ \nu(\mu_i) \right]^{-1} \frac{\partial \mu_i}{\partial \eta_i}
= \mu_i(1 - \mu_i) \left[ \frac{n}{\mu_i(1 - \mu_i)} \right] \mu_i(1 - \mu_i)
= n \mu_i(1 - \mu_i)
\]

\[
h_{ryi} = \frac{\partial \mu_i}{\partial \eta_i} \left[ \nu(\mu_i) \right]^{-1} \left[ y_i - \mu_i + \frac{\partial \mu_i}{\partial \eta_i} \eta_i \right]
= n \left[ y_i - \mu_i + \mu_i(1 - \mu_i) \eta_i \right].
\]

This can be translated into a Fortran subroutine as follows

```
SUBROUTINE LINK(Y,WT,ETA,MU,W,HRY)
REAL*8 Y,WT,ETA,MU,W,R,VAR,H
mu=exp(eta)/(1.+exp(eta)) ! h(\eta_i) = e^{\eta_i}/(1 + e^{\eta_i})
h=mu*(1.-mu) ! \partial \mu_i/\partial \eta_i = \mu_i(1 - \mu_i)
var=mu*(1.-mu)/wt ! \nu(\mu_i) = \mu_i(1 - \mu_i)/n
W=(H/VAR)*H ! \mathbf{W} = \text{Diag}(\mathbf{H}'\mathbf{R}^{-1}\mathbf{H})
HRY=(H/VAR)*(Y-MU)+W*ETA ! h_{ry} = \mathbf{H}'\mathbf{R}^{-1}[\mathbf{y} - \mu + \mathbf{H}\eta]
RETURN
END
```

In the subroutine \( Y = y_i \), \( WT = n \), and \( ETA = \hat{\eta}_i \) are passed to the subroutine. The subroutine returns \( W = W_{ii} \), \( MU = \mu_i \), and \( HRY = h_{ryi} \). In the subroutine the lines

```fortran
mu=exp(eta)/(1.+exp(eta))
h=mu*(1.-mu)
```

would need to be changed if a different link function was selected. The line

```fortran
var=h/wt
```

would need to be changed if a different variance function was selected. The changes to the `LINK` subroutine to take into account boundary conditions will be discussed in section 6.

For each round of solving the estimating equations (2), a new set of linear predictors needs to be calculated. This can be accomplished during the construction of the LHS and RHS assuming solutions are not destroyed. The Fortran code for this

```fortran
DO I=1,N ! Loop to read in the N records.
    Read in record
END
```

66
ETA=0
DO J=1,NEFF
   ETA=ETA+X*SOL
END DO
CALL LINK(Y,WT,ETA,MU,W,HRY)

Build LHS and RHS

4.2 Numerical example

The data in Table 2 have been constructed to illustrate the calculations involved for binomial random variables with a logit link function. A linear predictor of pregnancy rate for a cow in herd $h$ on diet $d$ is

$$
\mu_{hd} = \mu + \text{Herd}_h + \text{Diet}_d
$$

where $\text{Herd}_h$ is the effect of herd $h$ and $\text{Diet}_d$ is the effect of diet $d$. For the first round we will use an initial estimate for $\eta_i$ of 0. The contributions of each observation for round one is given in Table 3.

The estimating equations and solutions for round 1 are

$$
\begin{pmatrix}
37.5 & 12.5 & 25 & 15 & 20 & 2.5 \\
12.5 & 12.5 & 0 & 5 & 7.5 & 0 \\
25 & 0 & 25 & 10 & 12.5 & 2.5 \\
15 & 5 & 10 & 15 & 0 & 0 \\
20 & 7.5 & 12.5 & 0 & 20 & 0 \\
2.5 & 0 & 2.5 & 0 & 0 & 2.5
\end{pmatrix}
\begin{pmatrix}
\hat{\mu} \\
\hat{\text{Herd}}_1 \\
\hat{\text{Herd}}_2 \\
\hat{\text{Diet}}_A \\
\hat{\text{Diet}}_B \\
\hat{\text{Diet}}_C
\end{pmatrix}
= \begin{pmatrix}
40 \\
5 \\
35 \\
12 \\
25 \\
3
\end{pmatrix}
$$

and

$$
\hat{\eta} = X\hat{\beta} = \begin{pmatrix}
0.103896 \\
0.597403 \\
1.14805 \\
1.64156 \\
1.2
\end{pmatrix}
$$

The new estimates of the linear predictor are used to obtain the contributions of each observation for round two given in Table 4.
### Table 2: Example data.

<table>
<thead>
<tr>
<th>Herd</th>
<th>Diet</th>
<th>Number of Cows</th>
<th>Pregnancy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>20</td>
<td>0.50</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>30</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>40</td>
<td>0.80</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>50</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>10</td>
<td>0.80</td>
</tr>
</tbody>
</table>

### Table 3: Contributions to the estimating equations for round one.

<table>
<thead>
<tr>
<th>Herd</th>
<th>Diet</th>
<th>$y_i$</th>
<th>$\eta_i$</th>
<th>$\mu_i$</th>
<th>$n_i$</th>
<th>$W_{ii}$</th>
<th>$h_i y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>20</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>0.6</td>
<td>0</td>
<td>0.5</td>
<td>30</td>
<td>7.5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>0.8</td>
<td>0</td>
<td>0.5</td>
<td>40</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.9</td>
<td>0</td>
<td>0.5</td>
<td>50</td>
<td>12.5</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>0.8</td>
<td>0</td>
<td>0.5</td>
<td>10</td>
<td>2.5</td>
<td>3</td>
</tr>
</tbody>
</table>

### Table 4: Contributions to the estimating equations for round two.

<table>
<thead>
<tr>
<th>Herd</th>
<th>Diet</th>
<th>$y_i$</th>
<th>$\eta_i$</th>
<th>$\mu_i$</th>
<th>$n_i$</th>
<th>$W_{ii}$</th>
<th>$h_i y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>0.5</td>
<td>0.103896</td>
<td>0.525951</td>
<td>20</td>
<td>4.98653</td>
<td>-0.000932566</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>0.6</td>
<td>0.597403</td>
<td>0.645062</td>
<td>30</td>
<td>6.86871</td>
<td>4.75153</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>0.8</td>
<td>1.14805</td>
<td>0.759155</td>
<td>40</td>
<td>7.31355</td>
<td>10.0301</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.9</td>
<td>1.64156</td>
<td>0.837747</td>
<td>50</td>
<td>6.79635</td>
<td>14.2693</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>0.8</td>
<td>1.2</td>
<td>0.768525</td>
<td>10</td>
<td>1.77894</td>
<td>2.44949</td>
</tr>
</tbody>
</table>
5 Variance Component Estimation

The code given above assumes that the variance components are known. In this section I will discuss how estimates of the variance components can be obtained. Conceptually the variance component problem can be broken into two parts. The first part is the estimation of the variance components associated with the random effects. The second part is the estimation of the variance components associated with the error distribution. Before examining the modifications for a GLMM we will briefly review the major features of a variance component estimation program for a mixed model.

5.1 Mixed model

Derivative based programs for estimation of variance components under a mixed model involve the computation of quadratic forms of the predicted random effects, \( \hat{u}Q_i \hat{u} \), along with functions of the elements of a generalized inverse of the left hand sides, \( f_{ij}(C) \), where \( C \) is a generalized inverse of the left hand sides in (2). For example, the univariate Fisher scoring quadratic form of the REML estimator of variance component \( i \) is

\[
\frac{\hat{u}_i' I_q \hat{u}_i}{\sigma_i^4}
\]

where \( \hat{u}_i \) is the \( q_i \times 1 \) vector of predicted random effects for \( i \)th set of random effects. The function of the left hand sides are

\[
f_{ii}(C) = \frac{1}{\sigma_i^2} \left[ q_i - \frac{2 \text{tr}(C_{ii}^2)}{\sigma_i^2} + \frac{\text{tr}(C_{ii} C_{ii}^2)}{\sigma_i^4} \right]
\]

\[
f_{ij}(C) = \frac{1}{\sigma_i^2 \sigma_j^2} \left[ \frac{\text{tr}(C_{ij} C_{ji})}{\sigma_i^2 \sigma_j^2} \right]
\]

where

\[
C = \begin{pmatrix}
C_{00} & C_{01} & \ldots & C_{0r} \\
C_{10} & C_{11} & \ldots & C_{1r} \\
\vdots & & & \vdots \\
C_{r0} & C_{r1} & \ldots & C_{rr}
\end{pmatrix}
\]

is the partitioned generalized inverse of the left hand sides of (2).

For the residual variance component, the quadratic form is

\[
\frac{(y - X \hat{\beta} - Z \hat{u})' I_N(y - X \hat{\beta} - Z \hat{u})}{\sigma_0^4}
\]

where \( N \) is the number of observations. The functions of the left hand sides are

\[
f_{i0}(C) = \frac{1}{\sigma_i^2 \sigma_0^2} \left[ \frac{\text{tr}(C_{ii}^2)}{\sigma_i^2} - \sum_{j=1}^r \frac{\text{tr}(C_{ij} C_{ji})}{\sigma_i^2 \sigma_j^2} \right]
\]

\[
f_{00}(C) = \frac{1}{\sigma_0^4} \left[ N - p^* - q + \sum_{i=1}^r \sum_{j=1}^r \frac{\text{tr}(C_{ij} C_{ji})}{\sigma_i^2 \sigma_j^2} \right]
\]

(3)
5.2 Modifications for GLMM

The variance components can be estimated using the approximate REML quasi-likelihood

\[
ql(\beta, \sigma) = -\frac{1}{2} \ln |V| - \frac{1}{2} \ln |X' H' V^{-1} H X| - \frac{1}{2} (y^* - H X \beta)' V^{-1} (y^* - H X \beta)
\] (4)

where \(\sigma\) is the vector of variance component and \(V = R + H Z G Z' H'\). For the variance components associated with the random effects in \(G\) the estimating equations remain the same except \(\hat{u}\) and \(C\) are obtained using (2) instead of the usual mixed model equations.

Estimation of the variance components associated with the error distribution is more problematic. The quadratic form becomes

\[
(y - \mu)' R^{-1} \frac{\partial R}{\partial \phi} R^{-1} (y - \mu).
\] (5)

However, the corresponding functions for the left hand side in (3) for the linear mixed model assumes that \(R = I \sigma^2_\phi\). The functions of the left hand sides for \(\phi\) are

\[
f_{00}(C) = \left[ \text{tr}(\Phi) - 2 \text{tr}(\Omega) + \text{tr}(\Omega \Phi \Omega) \right]
\]

\[
f_{i0}(C) = \left[ \text{tr}(C^i (X' H' H X X' H' \Phi H Z H' \Phi H Z C^i)) \right]
\]

where \(C^i = (C^{i0} \quad C^{i1} \ldots C^{ir})\),

\[
\Phi = R^{-1} \frac{\partial R}{\partial \phi} R^{-1}
\]

and

\[
\Omega = (H X \quad H Z) C \begin{pmatrix} X' H' \\ Z' H' \end{pmatrix}.
\]

6 Some Computational Issues

While the mathematics are “straight forward,” implementation in practice is often challenging. Many of the difficulties associated with generalized linear mixed models are related to either estimates going to infinity or divide by zero problems. Consider the univariate logit model. The calculations involve \(\frac{1}{\mu(1-\mu)}\). Provided \(0 < \mu < 1\) this quantity is well defined. If \(\mu\) approaches either zero or one, then the quantity \(\frac{1}{\mu(1-\mu)}\) approaches infinity. Estimates of \(\mu\) of zero or one occur when a contemporary group consists entirely of zeros or ones.

Several options exist for handling this situation. One approach is to remove from the data any troublesome contemporary groups. While mathematically sound, an additional edit is needed to remove legitimate data values. A second approach is to treat contemporary group effects as random effects. However, the decision to treat a set of effects as
fixed or random should be decided from a modeling standpoint and not as an artifact of the estimation procedure. A third approach is to adjust $y_i$ away from 0 and 1. For example, one could use $(y_i + \Delta)/(1 + 2\Delta)$ instead of $y_i$. A fourth approach is based on the examination of what happens to the quantities $W_{ii}$ and $h_{ri}$ when $\eta_i$ approaches infinity.

As $\eta_i \to \pm\infty$,

$W_{ii} \to 0$ and

$h_{ri} \to n(y_i - \mu_i)$

In the limit $W_{ii}$ and $h_{ri}$ are both well defined. One could then recode the link function as

```
SUBROUTINE LINK(Y,WT,ETA,MU,W,HRY)
REAL*8 Y,WT,ETA,MU,W,R,VAR,H
if(abs(eta) > BIG) then ! Check for $\hat{\eta}_i \to \pm\infty$
   if(eta > 0) then ! $\hat{\eta}_i \to \infty$ \Rightarrow $\hat{\mu}_i \to 1$
      mu=1.
   else ! $\hat{\eta}_i \to -\infty$ \Rightarrow $\hat{\mu}_i \to 0$
      mu=0
   end if
   w=0 ! $\hat{\eta}_i \to \pm\infty$ \Rightarrow $W_{ii} \to 0$
   hry=wt*(y-mu) ! and $h_{ri} \to n(y_i - \hat{\mu}_i)$
   return
mu=exp(eta)/(1.+exp(eta))
   h=mu*(1.-mu)
   var=h/wt
   w=(H/VAR)*H
   HRY=(H/VAR)*(Y-MU)+W*ETA
RETURN
END
```

where BIG is a sufficiently large, but not too large a number. For example BIG = 10 might be a reasonable choice. However, using $W_{ii} = 0$ will usually result in a zero diagonal in the estimating equations. When these equations are solved, the fixed effect which was approaching infinity will be set to zero. This substitution can result in interesting convergence problems. A solution would be to “fix” the offensive fixed effect by making the following changes

```
SUBROUTINE LINK(Y,WT,ETA,MU,W,HRY)
REAL*8 Y,WT,ETA,MU,W,R,VAR,H
if(abs(eta) > BIG) then ! Check for $\hat{\eta}_i \to \pm\infty$
   if(eta > 0) then ! $-\text{BIG} \leq \hat{\eta}_i \leq \text{BIG}$
      eta=BIG
   else ! $-\text{BIG} \leq \hat{\eta}_i \leq \text{BIG}$
      eta=-BIG
   end if
```
end if
mu=exp(eta)/(1.+exp(eta))
h=mu*(1.-mu)
var=h/wt
w=(H/VAR)*H
HRY=(H/VAR)*(Y-MU)+W*ETA
RETURN
END

6.1 Additional Parameters

Threshold models with more than two classes and growth curve models require more than one linear predictor per observation. In the case of a threshold model the value of an observation is determined by the usual linear predictor and a threshold linear predictor. With growth curve models animal i has n_i observations. The linear predictor for animal i is a p × 1 vector, where p is the the number of growth curve parameters.

Unlike the univariate case, the matrices H, W, and R may not be diagonal. Typical structures for these matrices are:

\[ R = \text{Diag}(R_i) \text{ with } R_i \text{ a } n_i \times n_i \text{ residual covariance matrix} \]
\[ H = \text{Diag}(H_i) \text{ with } H_i \text{ a } n_i \times p \text{ matrix of partial derivatives} \]
\[ H_i = \begin{pmatrix}
\frac{\partial \mu_{i1}}{\partial \eta_{i1}} & \cdots & \frac{\partial \mu_{i1}}{\partial \eta_{ip}} \\
\vdots & \ddots & \vdots \\
\frac{\partial \mu_{ni}}{\partial \eta_{i1}} & \cdots & \frac{\partial \mu_{ni}}{\partial \eta_{ip}}
\end{pmatrix} \]
\[ W = \text{Diag}(W_i) \text{ with } W_i = H_i' R_i^{-1} H_i. \]

7 Conclusions

Generalized linear mixed models provide a flexible way to model production traits which do not satisfy the assumptions of a linear mixed model. This flexibility allows the researcher to focus more on selecting an appropriate model as opposed to finding manipulations to make the data fit a restricted class of models. The flexibility of a generalized linear mixed model provides an extra challenge when selecting an appropriate model. As a general rule the aim should be to select as simple a model as possible which does a reasonable job of modeling the data.

While generalized linear mixed model programs are not as readily available as linear mixed model programs, modifications needed for a linear mixed model program should be minor. The two changes that are needed are to add a Link subroutine and to solve iteratively the estimating equations. In constructing a link subroutine it is important to handle boundary conditions robustly.

When selecting a link function it is important to remember that when you get the results from the analysis, you will need to both interpret the results and to present the results in a meaningful manner. Generalized linear mixed models provide us with a very powerful
tool.

References


