Generalized Linear Mixed Models

- Extend the generalized linear model to include both fixed effects and random effects.
- This is still a fairly new class of models.

An example of this would arise in a clinical study. Suppose that we are interested in testing a drug in humans. The response that we are interested in is the incidence of nausea in the subjects. We give the different individuals each of several drugs (including placebo) and we ask them if they felt nauseous during the morning following the drug treatment.

Here, the responses are either a 0 (no nausea) or a 1 (nausea). Thus, we would like to have a logistic regression model to predict the probability of nausea for a particular drug. However, some people are much more susceptible to nausea than others, and thus responses on an individual may be correlated.

To develop the generalized linear mixed model, we should consider the following. Recall that for the linear mixed model, we specified \( f(y|\alpha) \) and \( f(\alpha) \), and thus also \( f(y) \).

For linear mixed models, we assumed that all three of these distributions were linear and gaussian. Suppose now that we relax the normality assumption for \( f(y|\alpha) \). Rather, let’s assume

\[
y_i|\alpha \sim \text{indep}. f(y_i|\alpha),
\]
Generalized Linear Mixed Models, Cont.

and $f(y|\alpha)$ is an exponential family distribution,

$$f(y|\alpha) = \exp \left[ \{y_i\theta_i - b(\theta_i)\}/a(\phi) - c(y_i, \phi) \right].$$

The conditional mean of $y_i$ is related to $\theta_i$ according to the model

$$\mu_i = \frac{\partial b(\theta_i)}{\partial \theta_i}.$$

It is a transformation of this mean that we desire to model as a linear model in both the fixed and random effects. Let

$$E(y_i|\alpha) = \mu_i \quad \text{and} \quad g(\mu_i) = X'_i\beta + Z'_i\alpha,$$

where $g()$ is a known link function. Notice the similarity in notation to that for GLM. Here, we have replaced $b$ with $\alpha$ and let $\mu_i$ be the conditional mean.

Finally, to complete the specification of this model, let

$$\alpha \sim f(\alpha),$$

be some distribution.

**Consequences of having random effects:**

Recall that the mean of $y$ has the form

$$E(y_i) = E_\alpha\{E(y_i|\alpha)\} = E_\alpha(\mu_i) = E\{g^{-1}(X'_i\beta + Z'_i\alpha)\}.$$

In general, the form cannot be simplified because the form $g^{-1}(\cdot)$ is nonlinear.
Consequences of Random Effects, Cont.

In special cases, however, this expectation can be worked out. For example, consider a log link, \( g(\mu) = \log \mu \), or equivalently, \( g^{-1}(\cdot) = \exp(\cdot) \). Then,

\[
E(y_i) = E\{\exp(X'_i \beta + Z'_i \alpha)\} = \exp(X'_i \beta)E\{\exp(Z'_i \alpha)\}.
\]

This expectation is the moment generating function of \( \alpha \) evaluated at \( Z_i \).

Now, suppose that \( \alpha_i \sim \text{iid } N(0, \sigma^2_\alpha) \) and that each of the rows of \( Z \) has a single entry equal to 1, with all of the rest equal to 0. Then,

\[
E\{\exp(Z'_i \alpha)\} = \exp(\sigma^2_\alpha / 2) \quad \text{and} \quad E(y_i) = \exp(X'_i \beta) \exp(\sigma^2_\alpha / 2).
\]

We could equivalently write this as \( \log E(y_i) = X'_i \beta + \sigma^2_\alpha / 2 \).

**Variance of \( y \):**

First, recall that

\[
\text{var}(y_i) = \text{var}_\alpha\{E(y_i | \alpha)\} + E_\alpha\{\text{var}(y_i | \alpha)\}
\]

\[
= \text{var}(\mu_i) + E\{a(\phi)V(\mu)\}
\]

\[
= \text{var}\{g^{-1}(X'_i \beta + Z'_i \alpha)\} + E[a(\phi)V\{g^{-1}(X'_i \beta + Z'_i \alpha)\}].
\]

Once again, this form cannot be simplified in general without specific assumptions about \( g(\cdot) \) and/or about this distribution.
Consequences of Random Effects, Cont.

Covariance of $y$:

As in the case of a linear mixed model, the presence of random effects introduces a correlation among observations which share any random effect in common. Here,

$$\text{cov}(y_i, y_j) = \text{cov}\{E(y_i|\alpha), E(y_j|\alpha)\} + E\{\text{cov}(y_i, y_j|\alpha)\}$$

$$= \text{cov}(\mu_i, \mu_j) + E(0)$$

$$= \text{cov}\{g^{-1}(X_i\beta + Z_i\alpha), g^{-1}(X_j\beta + Z_j\alpha)\}.$$  

Here, we notice the importance of conditioning to induce the covariability.

**Example:**

Consider the problem of modelling data in correlated “clusters”, which are thought to come from a Poisson distribution. For example, Diggle, et al (1994) consider the number of epileptic seizures in patients on a drug or placebo with repeated measurements on the same patients. Let

$$y_{ij}$$ be the $j$th count taken on the $i$th patient. 

Then, let the model be $y_{ij}|\alpha \sim \text{indep. Pois}(\mu_{ij})$. Here,

$$\log(\mu_{ij}) = X'_{ij}\beta + \alpha_i,$$ where $\alpha_i \sim \text{iid N}(0, \sigma^2_\alpha).$

Note that in this example, we are using a log-link with a random patient effect.
Estimation for GLMMs

Likelihood:

It is easy in principal to write down the likelihood:

\[ L = \int f(y|\alpha) f(\alpha) d\alpha, \]

a \( q \)-dimensional integral, since \( \alpha \) has dimension \( q \).

However, in nearly all cases, this cannot be evaluated in a closed form. In some cases, for example, low dimensional \( \alpha \), standard numerical integration approaches can be considered. Then, we must consider \( \frac{\partial \ell}{\partial \beta} \), which can be shown to have the form

\[ \frac{\partial \ell}{\partial \beta} = X'E(W^*y) - X'E(W^*\mu|y), \]

where

\[ W^* = \text{diag}[\{a(\phi)V(\mu_i)g(\mu_i)\}^{-1}]. \]

The likelihood equation for \( \beta \) is then

\[ X'E(W^*y) = X'E(W^*\mu|y). \]

In some cases, for example, the Poisson, \( W^* = I \), so the likelihood equation is

\[ X'y = X'E(\mu|y). \]

Even this equation must be solved numerically in general.

For random effects, we could consider the corresponding forms for \( \frac{\partial \ell}{\partial \phi} \).
Generalized Estimating Equations

Let’s propose a marginal generalized linear model for the mean of \( y \) as a function of the predictors. For example, for binary data, we might have the model

\[
\text{logit}\{E(y)\} = X\beta.
\]

If we assume some working covariance matrix, \( V \), for the elements of \( y \), then the maximum likelihood equations for estimating \( \beta \) would be

\[
X'V^{-1}y = X'V^{-1}E(y).
\]

When \( V \) is correct, these are unbiased estimating equations. Often, we define \( V = I \).

Operationally, we could assume some covariance structure, and then do a standard logistic regression. Then, we could properly calculate its large sample variance. Here, the estimates are consistent.

Let \( Y_{ij}, \ j = 1, \cdots, n_i, \ i = 1, \cdots, K \) be the \( j \)th measurement on the \( i \)th subject. Let

\[
Y_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix} \quad \text{with mean } \mu_i = \begin{pmatrix} \mu_{i1} \\ \vdots \\ \mu_{in_i} \end{pmatrix},
\]

and let \( X_{ij} = \begin{pmatrix} X_{ij1} \\ \vdots \\ X_{ijp} \end{pmatrix} \).

Also, let \( V_i = \text{cov}(Y_i) \).
Generalized Estimating Equations, Cont.

Then, GEE (according to Liang and Zeger, 1986) estimates for $\beta$ have the form

$$S(\beta) = \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V^{-1} \{Y_i - \mu_i(\beta)\} = 0.$$ 

Now, since $g(\mu_{ij}) = X_{ij}' \beta$,

$$\frac{\partial \mu_i'}{\partial \beta} = \begin{bmatrix} \frac{X_{i11}}{g'(\mu_{i1})} & \cdots & \frac{X_{i1n_{ij}}}{g'(\mu_{i1})} \\ \vdots & \ddots & \vdots \\ \frac{X_{in_{ij}}}{g'(\mu_{i1})} & \cdots & \frac{X_{in_{ij}}}{g'(\mu_{i1})} \end{bmatrix}.$$ 

Let $R_i(a)$ be an $n_i \times n_i$ "working" correlation matrix specified up to some parameters $a$. Then, $V_i = \phi A_i^{1/2} R(a) A_i^{1/2}$, where $A_i$ is an $n_i \times n_i$ diagonal matrix with $V(\mu_{ij})$ on the $j$th diagonal. If $R(a)$ is the true correlation matrix of $Y_i$, then $V_i$ is the true covariance matrix.

Usually, the working correlation matrix must be estimated iteratively. We can do this according to the fitting algorithm.

1. Compute the initial estimate of $\beta$ (often using GLM under the independence assumption).

2. Compute the working correlation matrix $R$ based upon the studentized residuals.
Generalized Estimating Equations, Cont.

3. Compute the estimated covariance, $\hat{V}_i$.

4. Update $\beta$ according to

$$\beta_{r+1} = \beta_r + \left( \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \times \left\{ \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i) \right\}$$

5. Iterate until the algorithm converges.

Note that the is in general not a likelihood estimator. Thus, we cannot make any inferences based upon the likelihood, as they would not be appropriate.

**Other Approaches:**

Some other approaches that we could consider include penalized quasi-likelihood, or conditional likelihood. Additionally, we could consider a Bayesian method.

We would begin by putting priors on our parameters

$$f(\alpha, \beta|y) \propto f(y|\alpha, \beta) f(\alpha) f(\beta).$$

Then, numerical techniques, such as Markov Chain Monte Carlo (MCMC) can be used to find the desired distribution. Under certain conditions, this can be the “best” method.
Fitting these models in practice

• Proc GENMOD in SAS can now fit GLMMs with covariance parameters using the repeated statement.

• The GLIMMIX macros in SAS is a very general procedure which uses the pseudo-likelihood approach.

• BUGS is a software package which will fit the Bayesian method.

• Other procedures continue to be developed.