GENERALIZED LINEAR MIXED MODELS: THEORY AND PRACTICE

1. Introduction and Definition

1.1. **Intuition.** In GLMM there are consequences for having random effects that we haven't seen before. Many of these consequences are related to the potentially nonlinear nature of the model via the link function.

1.2. **Structure of the GLMM.** We start with the conditional distribution of y given b. We assume conditional independence of the elements of y with each distribution belonging to the exponential family or similar.

$$y_i|b \sim indep.f_{Y_i|b}(y_i|b) \tag{1}$$

$$f(y_i, \theta_i, \tau^2, \omega_i) = \exp\left(\frac{\theta_i y_i - d(\theta_i)}{\tau^2}\omega_i\right) h(y_i, \tau^2, \omega_i) \tag{2}$$

As in the basic GLM case, we have a condition mean, μ_i, which is connected to a linear predictor, $\eta_i = x_i'\beta + z_i'b$, via an invertible link function g.

$$E[y_i|b] = \mu_i = \frac{\partial d(\theta_i)}{\partial \theta_i} = g^{-1}(\eta_i)$$

$$g(\mu_i) = \eta$$

We also need to specify the distribution of b. Most of the time we assume:

$$b \sim \mathcal{N}(\mathbf{0}, \Sigma\ (\theta))$$

where $\mathbf{0}$ is an n-dimensional vector of zeros and Σ is a q X q variance covariance matrix determined by parameter vector θ.

1.3. **Marginal properties.** All is the same as the GLM framework except μ_i represents the conditional mean instead of the marginal or unconditional mean. We need to derive the properties of the marginal so that we can better understand the assumptions made.

$$E[y_i] = E[E[y_i|b]] = E[\mu_i] = E[g^{-1}(x_i'\beta + z_i'b)]$$

We cannot continue if g is a non-linear function without specifying it. Let's suppose that $g(\mu) = log(\mu)$ and $g^{-1}(x) = \exp(x)$. So,

$$E[y_i] = E[\exp(x_i'\beta + z_i'b)] = \exp(x_i'\beta)E[z_i'b] = \exp(x_i'\beta)$$

Now we need to apply the assumption $b \sim \mathcal{N}(0, \sigma_b^2)$. Suppose that Z is some permutation of the identity matrix. $E[z_i'b]$ equals the moment generating function, $M_u(z_i)$, of b evaluated at z_i. We can then deduce that $M_u(z_i) = E[z_i'b]$ and $\eta_i = \log E[y_i] = x_i'\beta\ + \sigma_b^2/2$.

The marginal variance can be calculated in a similar manner:

$$\begin{aligned}\text{var}(y_i) =&\ \text{var}(E[y_i|b]) + E[\text{var}(y_i|b)] = \text{var}(\mu_i) + E[\tau^2 v(\mu_i)]\\ =&\ \text{var}(g^{-1}(x_i'\beta\ + z_i'b) + E[\tau^2 v(g^{-1}(x_i'\beta\ + z_i'b))]\end{aligned}$$

We can't continue without further specification of g or the conditional distribution of y. Let's suppose $y_i|b \sim Poisson(\mu_i)$ and the log link for g.

$$\text{var}(y_i) = \text{var}(\mu_i) + E[\mu_i] = \text{var}(\exp(x_i'\beta\ + z_i'b)) + E[\exp(x_i'\beta\ + z_i'b)]$$

$$= E(\exp(2(x_i'\beta\ + z_i'b)) - E[\exp(x_i'\beta\ + z_i'b)]^2 + E[\exp(x_i'\beta\ + z_i'b)]$$

$$= \exp(2(x_i'\beta\ ))(M_u(2z_i) - [M_u(z_i)]^2 + \exp(-x_i'\beta\ )M_u(z_i))$$

With a little algebra and the same normal assumptions on the random effects that we assumed before, we finally arrive at the following:

$$\text{var}(y_i) = E[y_i][\exp(x_i'\beta\ )(\exp(3\sigma_b^2/2) - \exp(\sigma_b^2/2)) + 1]$$

Notice that the term on the right is greater than one, which means that under our assumptions, our marginal variance will always be greater than the marginal expected value (it is "overdispersed"). If the conditional distribution of y_i given b is Poisson, then the marginal cannot be. Random effects are a way to attribute overdispersion to a particular source.

Marginal covariances and correlations can be derived using a similar process that we used for deriving the marginal variance.

2. Maximum Likelihood Estimation

2.1. **Conditional Likelihood.**

To find the maximum likelihood estimates, $\hat{\beta}$ and $\hat{\theta}$, we need to find the values of β and θ that maximize the following conditional likelihood:

$$f(y|\beta,\theta) = \int_b p(y|\beta,b)f(b|\Sigma(\theta))db \tag{3}$$

where $p(y|\beta,b)$ is the probability mass function of y given β and b, and $f(b|\Sigma(\theta))$ is the density at b usually assumed to be normal.

There is no guarantee that a closed form solution exists to this integral. The strategy employed in LME4 is to use a Laplace approximation. For the Laplace approximation, we will need to compute the conditional modes of the random effects. Conditional modes are calculated using Penalized Iterated Reweighted Least Squares (PIRLS).

2.2. **The Laplace Approximation.** The idea behind the Laplace approximation is to represent a "not nice" density function with a Gaussian. For simplicity, consider an example where we want to approximate density $p(x) = f(x)/z$ where $f(x) \geq 0$. First, take the taylor series expansion of $\log f(x)$.

$$\log f(x) = \log f(x_0) + \frac{\partial \log f(x)}{\partial x}|_{x=x_0} * (x - x_0) + \frac{1}{2}\frac{\partial^2 \log f(x)}{\partial x^2}|_{x=x_0} * (x - x_0)^2 + h.o.t. \tag{4}$$

The second term will be zero if $x_0 = x_{max}$. So the first step is to solve for x_{max}, which will yield the local maxima of the density.

Neglecting the higher order terms, evaluate (5) at x_{max} to get:

$$\log f(x) = \log f(x_{max}) + \frac{1}{2}\frac{\partial^2 \log f(x)}{\partial x^2}|_{x=x_{max}} * (x - x_0)^2 \tag{5}$$

$$\exp(\log f(x)) = \exp(\log f(x_{max}) + \frac{1}{2}\frac{\partial^2 \log f(x)}{\partial x^2}|_{x=x_{max}} * (x - x_0)^2) \tag{6}$$

$$\int \exp(\log f(x))dx = \exp(\log f(x_{max})) \int \exp\left[\frac{1}{2}\frac{\partial^2 \log f(x)}{\partial x^2}|_{x=x_{max}} * (x - x_0)^2)\right] dx \tag{7}$$

For notational simplicity let $\log f(x) = l(x)$. We can see the density function for a Gaussian if we let $\sigma^2 = -\frac{1}{l''(x_{max})}$.

$$\int e^{l(x)}dx \approx e^{l(x_{max})} \int \exp\left[-\frac{(x - x_{max})^2}{2\sigma^2}\right] dx \tag{8}$$

3 steps for the Laplace approximation: 1) Find local maximum of pdf f(x); 2) Calculate variance $\sigma^2 = -\frac{1}{f''(x_{max}}$; 3) Approximate the pdf with $p(x) \approx \mathcal{N}(x_{max}, \sigma^2)$.

LME4 uses the Laplace approximation to find the MLEs, $\hat{\beta}$ and $\hat{\theta}$, that maximize (4). It takes the log of the integrand of (4) and evaluates the second order taylor series at the conditional maximum $\tilde{b}(\beta, \theta)$ which is explained in the next section on PIRLS. On the deviance scale:

$$-2l(\beta, \theta|y) = -2\log(\int_b p(y|\beta, b)f(b, \Sigma(\theta)))db \tag{9}$$

$$\approx 2\log(\int_b \exp(-\frac{1}{2}[d(\beta, \tilde{b}, y) + \tilde{b}^{*T}\tilde{b}^* + b^T D^{-1}b])db) \tag{10}$$

$$= d(\beta, \tilde{b}, y) + \tilde{b}^{*T}\tilde{b}^* + \log|D| \tag{11}$$

where $d(\beta, \tilde{b}, y) = -2\log p(y|\beta, b)$ is the deviance function of the linear predictor and can be evaluated as the sum of deviance residuals.

2.3. **PIRLS.** In order to apply the Laplace approximation we need to compute for given β and θ the conditional modes of the random effects:

$$\tilde{b}(\beta, \theta) = argmax_b p(y|\beta, b)f(b, \Sigma(\theta)) \tag{12}$$

where β is incorporated as an offset, $X\beta$ is the contribution of the variance components and θ is part of a penalty term.

This will be very similar to the IRLS procedure used in the GLM case except that we are going to add a penalty term. We are going to take the minimum of $-\log p(y|\beta, b)$ instead of the maximum so that we can add our penalty.

$$\tilde{b}(\beta, \theta) = argmin_b \left[-\log p(y|\beta, b) + \frac{b^T \Sigma^{-1}(\theta) b}{2} \right] \tag{13}$$

The solution will be the iteration of the following solution:

$$(Z^T W^{(r)} Z + \Sigma^{-1}) b^{(r+1)} = Z^T W^{(r)} z^{(r)} \tag{14}$$

Stop the iteration using the convergence criteria $\frac{||\eta^{(r+1)} - \eta^{(r)}||}{||\eta^{(r)}||}$. The variance-covariance matrix of b is $\text{var}(b|\beta, \theta, y) \approx D = (Z^T W^{(r)} Z + \Sigma^{-1})^{-1}$.

2.4. **Maximum Quasi Likelihood.** We often don't know a prior the distribution of the response variable. Quasi likelihood relies on fewer assumptions and is a work-around for this problem. We want to preserve certain characteristics of the mean-variance relationship to generate workable estimators.

$$E\left[\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i}\right] = 0 \tag{15}$$

$$\text{var}\left(\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i}\right) = \frac{1}{\tau^2 v(\mu_i)} \tag{16}$$

We want a quantity for $\frac{\partial \log f_{Y_i}(y_i)}{\partial \mu_i}$ with these properties. It turns out:

$$q_i = \frac{y_i - \mu_i}{\tau^2 v(\mu)} \tag{17}$$

One could attempt to define $Q_i = \int_{y_i}^{\mu_i} \frac{y_i - t}{\tau^2 v(t)} dt$ but we actually can maximize it without knowing exactly what it is.

$$\frac{\partial}{\partial \beta} \sum Q_i = 0 \tag{18}$$

$$\sum \frac{y_i - \mu_i}{\tau^2 v(\mu_i) g_\mu(\mu_i)} x_i' = 0 \tag{19}$$

This can also be expressed as:

$$\frac{1}{\tau^2} X' W \Lambda (y - \mu) = 0 \tag{20}$$

This method maximizes the likelihood only using the mean-to-variance relationship. We could also include a penalty term and use a penalized quasi likelihood (PQL) technique. However, LME4 does not do this.