Chapter 4

Generalized linear mixed models (GLMMs)

4.1 Introduction

I again begin with an example. Korff et al. (1994) studied the effects of physicians' practice style in treating back pain and its influence on functional measures (disability score, activity limitation days, etc.), patient satisfaction (e.g., "After your visit with the doctor, you fully understood how to take care of your back problem") and cost. Forty-four primary care physicians in a large HMO were classified according to their practice style in treating back pain management (low, moderate or high frequency of prescription of pain medication and bed rest). An average of 24 patients per physician was followed for 2 years (1 month, 1 year and 2 year followups) after the indexed visit. I'll focus on two types of questions: (1) Does practice style influence function, satisfaction or cost? and (2) How much variability is there in physician outcomes within a practice style?

There are a number of outcomes in this study, with a variety of different distributions. Outcomes like the disability score (which was calculated as the average of 3 scales, each on a range from 0-10) are almost certainly approximately normal and statistically well behaved (though averaging the 3 scales might create interpretational difficulties). An outcome like number of days on which activity was limited by the back pain might more properly be treated as Poisson distributed and an outcome like whether or not the patient understood the intended care is binary. Each of these would require a different distributional assumption and (probably) a different form of regression, indicating the use of generalized linear models.

There is an additional statistical complication with this study: there are multiple patients per physician and multiple measures (for some of the outcomes) per patient. This data is thus clustered or hierarchical in nature with predictors specific to each level of the data structure. As examples: at the physician level we have practice style, at the patient level we have age and gender, and at the visit level we have time since the indexed visit. The clustered nature of the data set significantly impacts the statistical analysis since the data must be regarded as correlated. Further, to answer a question about variability in outcomes attributable to physician differences, we

will need to be able to describe the degree to which outcomes are the same within physician and different between physicians.

4.2 Basic idea

The idea behind generalized linear mixed models (GLMMs) is conceptually straightforward: incorporate random effects into the linear predictor portion of a generalized linear model. This simple change allows us to accommodate correlation in the context of a broad class of models for non-normally distributed data. Viewed another way, it is a convenient way to build multivariate distributions for non-normal data that can accommodate some flexibility in the structure of the association as well as a rich set of predictor variables. Next I consider a stylized example in a bit more detail

4.3 Example: Skin cancer

This example is patterned after Abu-Libdeh et al. (1990) in which a model is developed for studying subjects who were given selenium in an attempt to reduce skin cancer risk. Let Y_{it} represent the number of newly discovered basal cell carcinoma sites on person i at time t, where we envision subjects being checked yearly over a five year study. Suppose we consider three predictors: SEX (the sex of the subject), SEL (whether or not the subject was in the selenium treatment group), and SUN (a measure of sun exposure over the past year). We might hypothesize that Y_{it} follows a Poisson distribution, but we would expect that some subjects are much more susceptible to skin cancer than others. So a reasonable model might be

$$\begin{aligned} Y_{it}|\boldsymbol{\lambda} &\sim \text{ Poisson}(\lambda_{it}), \\ \log(\lambda_{it}) &= \beta_{0i} + \beta_1 \text{SEX}_i + \beta_2 \text{SEL}_i + \beta_3 \text{SUN}_{it}, \end{aligned} \tag{4.1}$$

to which we add a final specification, namely that the subject-specific intercepts, β_{0i}, follow a distribution across subjects, centered at the value β_0:

$$\beta_{0i} \sim \mathcal{N}(\beta_0, \sigma^2_{s,0}). \tag{4.2}$$

a. Covariances

This assumption of random intercepts induces a correlation between measurements Y_{it} and $Y_{it'}$. More precisely, and using (2.3) we have

$$
\begin{aligned}
&\mathrm{Cov}(Y_{it}, Y_{it'}) \\
&\quad = \mathrm{E}[\mathrm{Cov}(Y_{it}, Y_{it'}|\boldsymbol{\lambda}_i)] + \mathrm{Cov}(\lambda_{it}, \lambda_{it'}) \\
&\quad = 0 + \mathrm{Cov}(\lambda_{it}, \lambda_{it'}) \\
&\quad = \mathrm{Cov}(\exp[\beta_{0i} + \beta_1 \mathrm{SEX}_i + \beta_2 \mathrm{SEL}_i + \beta_3 \mathrm{SUN}_{it}], \\
&\qquad\qquad \times \exp[\beta_{0i} + \beta_1 \mathrm{SEX}_i + \beta_2 \mathrm{SEL}_i + \beta_3 \mathrm{SUN}_{it'}]) \\
&\quad = \exp(2\beta_1 \mathrm{SEX}_i + 2\beta_2 \mathrm{SEL}_i + \beta_3 \mathrm{SUN}_{it} + \beta_3 \mathrm{SUN}_{it'}) \\
&\qquad \times \left[\mathrm{Cov}\left(e^{\beta_{0i}}, e^{\beta_{0i}}\right)\right] \\
&\quad = \exp(2\beta_1 \mathrm{SEX}_i + 2\beta_2 \mathrm{SEL}_i + \beta_3 \mathrm{SUN}_{it} + \beta_3 \mathrm{SUN}_{it'}) \\
&\qquad \times \left(\mathrm{E}[e^{2\beta_{0i}}] - \mathrm{E}[e^{\beta_{0i}}]^2\right) \\
&\quad = \exp(2\beta_1 \mathrm{SEX}_i + 2\beta_2 \mathrm{SEL}_i + \beta_3 \mathrm{SUN}_{it} + \beta_3 \mathrm{SUN}_{it'}) \\
&\qquad \times \left(\exp(\sigma^2_{s,0})\{\exp(\sigma^2_{s,0}) - 1\}\right),
\end{aligned}
\tag{4.3}
$$

with the last equality following from the moment generating function of a normal distribution. We can see that the covariance is positive as long as $\sigma^2_{s,0}$ is greater than zero.

b. Random slopes

Other parameters in the model can be assigned a distribution. Consider assigning a distribution to the β_3 parameter as follows:

$$
\begin{aligned}
Y_{it}|\boldsymbol{\lambda} &\sim \text{ Poisson}(\lambda_{it}), \\
\log(\lambda_{it}) &= \beta_{0i} + \beta_1 \mathrm{SEX}_i + \beta_2 \mathrm{SEL}_i + \beta_{3i} \mathrm{SUN}_{it}, \\
\beta_{0i} &\sim \mathcal{N}(\beta_0, \sigma^2_{s,0}), \\
\beta_{3i} &\sim \mathcal{N}(\beta_3, \sigma^2_{s,3}),
\end{aligned}
\tag{4.4}
$$

and the β_{0i} and β_{3i} may be correlated.

What changes has this induced? In (4.1), β_3 is the same for all subjects and has the interpretation as the sun exposure effect, which is assumed common for all subjects. In (4.4), β_{3i} is the sun exposure effect for the ith subject, so we are allowing some individuals to be more sensitive to the sun than others. This can be more formally investigated by testing the hypothesis $\mathrm{H}_0 : \sigma^2_{s,3} = 0$. That is, if there is no variance in the subject specific slopes, then model (4.1) holds. This may make a useful precursor to an attempt to find subgroups of sensitive individuals.

c. Prediction

Another inferential goal might be prediction. That is, we might be interested in predicting which individuals are more sensitive to the sun (especially if we had already established that $\sigma^2_{s,3} > 0$). To do so, we would want to identify those individuals with extreme values of β_{3i}. Since β_{3i} is a random variable, we would want to derive predictions of the realized values.

d. Unequal variances

Suppose we assume a distribution on the β_2 term in model (4.1) instead: $\beta_{2i} \sim \mathcal{N}(\beta_2, \sigma^2_{s,2})$. If SEL is coded as 1 for the selenium treatment group and 0 for the control, then the β_{2i} term appears in the model equation for the treatment group but not for the control group. If the β_{2i} are assumed independent of the other random effects then the variance in the treatment group is being modeled as larger than in the control group. Note that this would not necessarily be true if the β_{2i} and β_{0i} were allowed to be correlated. This emphasizes the point that, in general, these models allow for (or insist on, depending on your point of view) unequal variances. This can occasionally be of benefit, for example, if the focus is on whether the treatment group caused an increase in variability.

4.4 Specifying GLMMs

Specifying generalized linear mixed models involves making decisions about four aspects of the problem:

1. What is the distribution of the data for fixed values of the predictors?
2. What aspect of the problem will be modeled?
3. What are the predictors to be included in the model?
4. Which categorical predictors will be assumed to have a distribution?

Except for the addition of the final question this is the same list as for generalized linear models. So the extension is a straightforward but consequential one.

4.5 A more general model

As illustrated by the examples in this chapter, generalized linear mixed models are typically constructed by incorporating random effects into the linear predictor of a conditionally independent exponential family model. I now formalize that notion with the following definition of a *generalized linear mixed model*:

$$
\begin{aligned}
Y_i|\mathbf{u} &\sim \text{indep. } f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}), \\
f_{Y_i|\mathbf{u}}(y_i|\mathbf{u}) &= \exp\{[y_i\gamma_i - b(\gamma_i)]/\tau^2 - c(y_i,\tau)\}, \\
\mathrm{E}[Y_i|\mathbf{u}] &= \mu_i, \\
g(\mu_i) &= \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u}, \\
\mathbf{u} &\sim f_U(u).
\end{aligned}
\tag{4.5}
$$

In this definition we see the usual ingredients of a generalized linear model. First, the distribution of Y_i from an exponential family (in this case the distribution is assumed to hold conditional on the random effects $\mathbf{u}$). Second, a link function, $g(\cdot)$ is applied to the conditional mean of Y_i given $\mathbf{u}$ to obtain the *conditional linear predictor*. Finally, the linear predictor is assumed to consist of two components, the fixed effects portion, described by $\mathbf{x}_i'\boldsymbol{\beta}$ and the random effects portion, $\mathbf{z}_i'\mathbf{u}$, for which a distribution is assigned to $\mathbf{u}$.

4.6 Inference in GLMMs

How might we go about fitting a model like (4.5)? Maximum likelihood or variants (like REML) based on normality assumptions are relatively standard for linear mixed models. For example, SAS PROC MIXED fits by using ML or REML. For many GLMs, maximum likelihood is also standard, for example, logistic regression or Poisson regression. What about GLMMs?

In general, evaluation of the likelihood can be quite difficult. Consider a mixed logistic regression model for binary data. The likelihood would take the form

$$\int \cdots \int \exp\left\{\sum_i Y_i(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u})\right\} \prod_i \{1 + \exp(\mathbf{x}_i'\boldsymbol{\beta} + \mathbf{z}_i'\mathbf{u})\}^{-1} \times dF(\mathbf{u}), \tag{4.6}$$

where the integration is of a dimension equal to the dimension of $\mathbf{u}$. For some cases, for example random intercepts, (4.6) reduces to a product of one-dimensional integrals and hence can be evaluated numerically. In general, however, this is not true and for more complicated examples, for example, crossed random effects, the dimensionality of the integration quickly becomes unmanageable by standard numeric means.

In simple cases that are amenable to numerical integration, inference using ML would proceed using the usual asymptotic approximations:

- ML estimates are asymptotically normal, with standard errors coming from second derivatives of the log likelihood.
- Tests would be based on the likelihood ratio test, comparing twice the negative of the loglikelihood for nested models. Alternatively, Wald tests could be formed.
- Best predicted values would be estimated by calculating the expected value of the random effect conditional on the data and plugging in ML or REML estimates for unknown parameters. In general, the conditional expected values cannot be evaluated in closed form either.
- Tests of whether variances of random effects are zero can be based on the likelihood ratio statistic. However, as with linear mixed models, the asymptotic distribution is not a simple chi-square distribution.

To elaborate on this last point, the situation is the same as linear mixed models: when testing if a single variance component is equal to zero, the large-sample distribution under H_0 is a 50:50 mixture of a χ^2_1 and 0.

The difficulty of likelihood inference opens the door for alternatives. Two popular alternatives are Generalized Estimating Equations (GEEs), which are mainly for longitudinal data, and penalized quasi-likelihood. These will be covered in more detail in Chapter 8.

4.7 Further notes

The model as specified by (4.5) assumes that all the correlation can be described by the random effects. In some cases it is likely that other sources of correlation may be present, for example, time-series correlation. See Chan and Ledolter (1995), Jorgensen et al. (1999) and Davis et al. (2000) for examples of this type of model.