Chapter 9

Conclusions

9.1 Introduction

In Chapter 1 I introduced the chestnut leaf blight example and presented some preliminary analyses. None of these were correct since they did not accommodate the correlation introduced by the donor and recipient clones. A feature of the leaf blight example is the presence of crossed random effects. This eliminates the ability to use simpler models and/or longitudinal data approaches like GEEs. I now return to the example and illustrate the Monte Carlo ML approaches of Chapter 7.

9.2 Chestnut leaf blight revisited

I focus on selected portions of the model; those interested in the full analysis can consult Cortesi et al. (2001). Recall that the model, given in (5.1) and (5.2), was for the binary outcome of transmission of the virus between two (fungal) individuals and that repeated measurement on clones as either donors or recipients of the virus might introduce correlation. There is no reason to expect that the ability of a particular clone to transmit the virus (donor effect) is the same as its susceptibility to the virus (recipient effect). However, it is possible that the donor effect of a particular clone is correlated with the recipient effect. Accordingly, we fit a bivariate normal random effect distribution for the joint donor and recipient random effects associated with a clone.

The model was fit using a bivariate version of the MCNR algorithm described in Section 7.3. Standard errors were estimated by a similar simulation scheme applied to the information matrix and the values of the log likelihood were estimated by simulation using the method of Geyer and Thompson (1992). The estimate of the covariance between the donor and recipient random effects was almost zero and contributed negligibly to the likelihood. Also, the fourth putatively contributing gene had no effects (main, asymmetry or even third order effects) and hence that gene and the covariance were dropped and the model was refit with independent donor and recipient random effects. For that model the estimate of the donor variance component was $\sigma_d^2 = 0.28$ and the recipient variance component was $\sigma_r^2 = 1.24$ with standard errors (respectively) of 0.09 and 0.34.

These were statistically significant, but of smaller magnitude that most of the fixed effects in the model. For example the effects of individual genes (other than the fourth) to inhibit transmission when there was a mismatch ranged from -1.5 to -5.4, quite large for a logit model.

9.3 Further work

GLMMs have gained popularity in practice but a number of areas remain to be developed beyond those already identified. Choice of link function has been addressed somewhat in the GLM literature by, for example, Pregibon (1980); Li and Duan (1989); Mallick and Gelfand (1994); Weisberg and Welsh (1994); Xie et al. (1997) and a natural question is whether those techniques carry over directly to GLMMs.

a. Random effects distribution

The issue of sensitivity to assumptions of the random effects distribution is not completely resolved. Some literature (e.g., Neuhaus et al., 1992) indicates that the choice is not so crucial; others indicate that it may be (Butler and Louis, 1992; McCulloch, 1997; Piepho and McCulloch, 1999). This last reference proposes a model-based method for checking random effects distributions. Most likely the sensitivity to the choice of the distribution will manifest itself in inferences concerning the variances, the predicted values and quantities depending on the distributional shape, as opposed to the marginal mean. For an example of this see Magder and Zeger (1996).

It would be nice to have easily interpreted (perhaps graphical) tools for assessing the random effects distribution. Plotting of the best predicted values (e.g., histograms or Q-Q plots) has been suggested (Lange and Ryan, 1989), but the work of Verbeke and Lesaffre (1996) casts doubt on the ability of that method to discern distributional shape.

The matched pairs logit model of Chapter 6 is informative. Recall the model, given by

$$\begin{aligned} Y_{ij}|\alpha_i &\sim \text{indep. Bernoulli}(p_{ij}), \\ \text{logit}(p_{ij}) &= \alpha_i + \beta x_{ij}, \end{aligned} \tag{9.1}$$

where x_{ij} is 0 for the control and 1 for the treatment and the α_i are the pair effects. I use this to make two points.

First, with matched pairs and binary data, there are only four possible outcomes for each pair (two failures, two successes and two mixed results). A saturated model would therefore be a four category multinomial model. Clearly we will not be able to distinguish from the data any two mixing distributions that give the same four probabilities for the four categories. Essentially, any mixing distribution that is flexible enough to allow a wide range of these four probabilities will fit the data equally well. Hence results based on the estimated values of those probabilities will be similar across a variety of mixing distributions. This idea is developed more carefully in Neuhaus et al. (1992) and Neuhaus et al. (1994).

Secondly, a histogram or any other distributional assessment based on the $\tilde{\alpha}_i = \text{E}[\alpha_i|\mathbf{Y}]$ will be virtually non-informative since $\tilde{\alpha}_i$ can take on only four values

(corresponding to the four possible patterns of (Y_{i1}, Y_{i2}) pairs). Even if the true underlying distribution is normal, the distribution of the $\tilde{\alpha}_i$ will have at most four support points.

b. Small sample distributions

Small sample adjustments for classes of models simpler than GLMMs have been topics of recent interest, for example, linear mixed models (Kenward and Roger, 1997; Lyons and Peters, 2000) and generalized linear models (Kauermann and Carroll, 2001). Similar work is badly needed for GLMMs, where the issue of effective sample size is complicated by both the non-normality and the random effects. Approaches such as Fraser et al. (1999) appear promising.

With regard to inference for variance components even the large sample distribution theory is difficult to work out (Crainiceanu and Ruppert, 2002) and there is evidence that, in many cases, it fails to be a good approximation to small sample results (Crainiceanu and Ruppert, 2002; Pinheiro and Bates, 1995). It is enough to make even a dyed-in-the-wool frequentist turn to the dark side.

c. Prediction error

In the case of linear mixed models there is some work on the effect of plugging in estimates to best predicted values and its influence on standard errors (Peixoto and Harville, 1986; Harville, 1985; Jeske and Harville, 1988), but little for GLMMs (but see Have and Localio, 1999, for an exception). There is even a question as to the proper context for the calculation of prediction error (Booth and Hobert, 1998).

Typical interval estimates are calculated unconditional on $\mathbf{Y}$. But best predicted values are estimated conditional means, conditional on the observed values of $\mathbf{Y}$. Should the prediction error also be conditional, as argued in Booth and Hobert (1998)?

Consider a more concrete example. Suppose you must chose a hospital from which to get gall bladder surgery. You happen to live in a state that publishes sophisticated hospital profile reports, based on the predicted random effects for each hospital with regard to lack of complications at gall bladder surgery. Do you want the interval estimates of each of those predicted values to be conditional on the observed data that generated them? Or do you envision a broader inferential framework (unconditional on $\mathbf{Y}$) in which repeated samples are taken of patients, but best predicted values are derived for the same hospitals?

9.4 Summary

In summary, GLMMs have become heavily used due to the need for ways to model correlated, non-normally distributed data. Software is starting to mature and more experience has been gained in the utility and pitfalls of many of the techniques for analyzing GLMMs. Yet much further work remains to be performed before this arena can be considered a mature and thoroughly tested one.