15

Generalized Linear Mixed Effects Models: Approximate Methods of Estimation

15.1 INTRODUCTION

In the previous chapter we reviewed the broad class of regression models known as *generalized linear mixed effects models* (GLMMs) and described how they extend the conceptual approach of the linear mixed effects models to generalized linear models. By allowing a subset of the regression parameters in generalized linear models to vary from individual to individual, via the introduction of random effects, GLMMs account for natural heterogeneity in the study population. However, although GLMMs are conceptually straightforward to develop, model fitting and parameter estimation can be challenging because the likelihood to be maximized does not have a simple closed-form expression. Because the likelihood to be maximized is obtained by integrating or averaging over the distribution of the random effects, maximizing that likelihood is challenging.

In Chapter 14 we discussed ML estimation of the fixed effects and variance components in GLMMs through the use of adaptive Gaussian quadrature. This approach to estimation and inference can be considered "true" ML to the extent that an adequate number of quadrature points have been used to ensure good numerical accuracy. Recall that Gaussian quadrature approximates the likelihood to be maximized by replacing the integrals in the likelihood with summations. This approximation can be made very accurate by increasing the number of quadrature points. Of course, increasing the number of quadrature points raises the computational burden. For example, a tripling of the number of quadrature points for a GLMM with two random effects will lead to an almost 10-fold increase in computation.

Applied Longitudinal Analysis, Second Edition. By Garrett M. Fitzmaurice, Nan M. Laird, and James H. Ware.

Methods of estimation and inference that rely on numerical quadrature can become computationally burdensome as the number of random effects in the GLMM increases. For example, while adaptive Gaussian quadrature is relatively straightforward in a GLMM with only two random effects, it can be quite hopeless in the setting of a GLMM with more than 10 random effects where it breaks down due to the "curse of dimensionality." To see why, note that 30-point quadrature in a GLMM with two random effects requires numerical evaluations at a total of 900 (or 30^2) quadrature points. In contrast, 30-point quadrature in a GLMM with 10 random effects requires numerical evaluations at over 590 trillion (or 30^{10}) quadrature points! The computations grow exponentially with the number of random effects. This has provided the impetus for the development of alternative approximations that are far less computationally demanding. These alternative methods of estimation and inference for GLMMs are the topic of this chapter. Specifically, we review two approximate methods of estimation known as *penalized quasi-likelihood* (PQL) and *marginal quasi-likelihood* (MQL). Both of these approximate methods have been implemented in various statistical software packages (e.g., PROC GLIMMIX in SAS) and stand-alone programs that have been specifically tailored for fitting GLMMs (e.g., MLwiN). However, as we will see later, the PQL method is the only *legitimate* approximate method in the sense of yielding estimates of the parameters of the GLMM. In contrast, the MQL method, although derived as an approximate method of estimation for GLMMs, does not actually yield estimates of the parameters of the GLMM. This critical distinction between the PQL and MQL methods is one that is not well recognized; in Section 15.4 we highlight the important differences between the methods and their consequences for inference.

Finally, we emphasize at the outset that in many longitudinal studies there is only a single level of nesting in the data (i.e., repeated occasions nested within individuals) and the dimension of the vector of random effects (e.g., q) is relatively small (e.g., $q \leq 3$). Moreover, with longitudinal binary data, there is usually not much information available about random effects beyond a single variance component, especially when the number of repeated measurements is relatively small. Thus in the longitudinal setting the need for these alternative approximate methods is somewhat less acute than in other settings with cluster-correlated data where there may be numerous levels of nesting in the data and where the dimensionality of the random effects may be relatively large. This is important because many of the alternative approximate methods discussed in this chapter can perform poorly in certain settings relative to adaptive Gaussian quadrature. In addition the use of PQL and MQL methods is even more problematic when longitudinal data are incomplete. The inaccuracy of their approximatations implies that the methods are not valid when data are missing at random (MAR), but not missing completely at random (MCAR); see Section 4.3 for the definitions of, and the distinction between, MCAR and MAR. As a result these methods can produce badly biased estimates of the effects of covariates when data are MAR, but not MCAR. Therefore we consider ML estimation based on adaptive Gaussian quadrature, with a sufficiently large number of quadrature points, to be the method of choice for fitting GLMMs to longitudinal data.

15.2 PENALIZED QUASI-LIKELIHOOD

Penalized quasi-likelihood (PQL) estimation of GLMMs has been proposed as a computationally simple alternative to methods based on numerical quadrature, especially when the number of random effects is relatively large. There are two closely related derivations of the PQL algorithm. The first is obtained by applying an approximation to the integrand (i.e., the function to be integrated or averaged over) in the GLMM likelihood so that a closed-form expression for the integral is achieved. The particular approximation used is known as a conventional (or lower-order) Laplace approximation. Interestingly there is a connection between this Laplace approximation and Gaussian quadrature. It can be shown that a Laplace approximation to the integrand in the likelihood for a GLMM corresponds to the use of adaptive Gaussian quadrature with only a single quadrature point. The Laplace approximation is accurate in cases where the conditional distribution of the vector of responses, given the random effects, is approximately normal but can provide a poor approximation otherwise. Later we discuss settings where the approximation is likely to be accurate versus settings where it may yield badly biased estimates. The use of a conventional Laplace approximation is one motivation for PQL.

The second, and closely related, derivation for PQL follows from an approximation to the GLMM. Recall from Chapter 14 that the GLMM can be expressed as

$$g\{E(Y_{ij}|b_i)\} = \eta_{ij}^b = X_{ij}'\beta + Z_{ij}'b_i \tag{15.1}$$

for some known link function, $g(\cdot)$, and a $q \times 1$ vector of random effects $b_i \sim N(0, G)$. In a very slight departure from the notation used in Chapter 14, the superscript "b" in η_{ij}^b (and, later, in μ_{ij}^b) is used to emphasize conditioning on the random effects, b_i. Because the conditional distribution of each Y_{ij}, given b_i, belongs to the exponential family of distributions (e.g., Bernoulli or Poisson), $\text{Var}(Y_{ij}|b_i) = v\{E(Y_{ij}|b_i)\}\,\phi$, where $v(\cdot)$ is a known variance function. Note that the model for the conditional mean given by (15.1) can also be expressed as:

$$Y_{ij} = g^{-1}(X_{ij}'\beta + Z_{ij}'b_i) + \epsilon_{ij} = \mu_{ij}^b + \epsilon_{ij}, \tag{15.2}$$

where $g^{-1}(\cdot)$ is the *inverse* link function, μ_{ij}^b denotes the *conditional* mean of Y_{ij}, given b_i, and the errors ϵ_{ij} are assumed to have a mean of zero. The PQL method proceeds by approximating the model given by (15.2), so that a linear mixed effects model holds, at least approximately, for a transformation of the response (and a related transformation of the errors); the transformation of the response is denoted by Y_{ij}^*. Specifically, the standard linear mixed effects model,

$$Y_{ij}^* \approx X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij}^*, \tag{15.3}$$

holds approximately for the "working" response (or so-called pseudo-data), Y_{ij}^*. The "working" response Y_{ij}^*, a transformation of Y_{ij}, depends on current estimates of the fixed effects, β, and the random effects, b_i; similarly the transformed error, ϵ_{ij}^*, depends on current estimates of β and b_i. Specifically, when a canonical link function

has been chosen,

$$Y_{ij}^* = v^{-1}(\widehat{\mu}_{ij}^b)(Y_{ij} - \widehat{\mu}_{ij}^b) + X_{ij}'\widehat{\beta} + Z_{ij}'\widehat{b}_i$$

and $\epsilon_{ij}^* = v^{-1}(\widehat{\mu}_{ij}^b)\epsilon_{ij}$. Recall from Section 11.2 that canonical link functions are simply unique transformations of the mean that can be derived for any selected distribution (e.g., logit link function for a Bernoulli response or log link function for a Poisson response).

Although the technical details behind this approximation are not important, we note that (15.3) is referred to as a first-order Taylor series expansion of (15.2) around current estimates $\widehat{\beta}$ and $\widehat{b}_i$. A very similar type of expression for Y_{ij}^* can be derived when a non-canonical link function is adopted. A brief derivation of the approximation, together with an illustration of the form of the "working" response Y_{ij}^* for a mixed effects logistic regression model, can be found in Section 15.7; readers who find the level of detail in this section challenging can omit Section 15.7 at first reading without loss of continuity.

Recognizing that the "working" response Y_{ij}^* follows a linear mixed effects model, with fixed effects β, and with random effects b_i and within-subject errors ϵ_{ij}^*, estimation can proceed by iterating between the following two steps:

1. Fit a linear mixed effects model to the "working" response Y_{ij}^*, to obtain updated estimates of β and G, and subsequently, empirical BLUP predictions of b_i. The linear mixed effects model is fitted using weights, w_{ij}, that are inversely proportional to the variance of the ϵ_{ij}^*; for example, $w_{ij} = \phi^{-1}\, v(\widehat{\mu}_{ij}^b)$ when a canonical link function is adopted.

2. Use the estimates of β and b_i obtained in step 1 to update the "working" response Y_{ij}^* (and also update the weights, w_{ij}, used for estimation in step 1).

This two-step algorithm can be iterated until convergence has been achieved. The resulting estimates are known as the PQL estimates. Recall that in the linear mixed effects model the estimator of β is the *generalized least squares* (GLS) estimator that can be expressed as

$$\widehat{\beta} = \left\{ \sum_{i=1}^{N} \left(X_i' V_i^{*-1} X_i \right) \right\}^{-1} \sum_{i=1}^{N} \left(X_i' V_i^{*-1} Y_i^* \right), \tag{15.4}$$

where Y_i^* is the $n_i \times 1$ vector of "working" responses (with components Y_{ij}^*), X_i is the corresponding matrix of covariates, $V_i^* = Z_i G Z_i' + W_i^{-1}$ is the marginal covariance of Y_i^*, Z_i is the matrix of covariates for the random effects, and W_i is a diagonal weight matrix (with components w_{ij} along the diagonal). Estimation of the variances (and covariances) of the random effects can be based on the standard ML or REML estimators for a linear mixed effects model applied to the "pseudo-data". This method of estimation for the variances (and covariances) of the random effects is often referred to as *pseudo-likelihood* (PL).

There are two important points to recognize about the PQL method. First, PQL is an approximate method. This means that in some settings the approximation is quite

accurate and produces valid estimates of the fixed effects (and covariance parameters of the random effects), while in other settings it is not accurate and can yield badly biased estimates. In Section 15.4, we consider the factors that influence the accuracy of the approximation. Second, the PQL method produces an estimate of the fixed effects, β, in the GLMM:

$$g\{E(Y_{ij}|b_i)\} = X'_{ij}\beta + Z'_{ij}b_i,$$

albeit with varying degrees of bias that depend on the accuracy of the approximation. Thus, the PQL method is appropriate when the goal of the analysis is to make subject-specific inferences.

Finally, we note that there can be small differences in the implementation of the PQL algorithm in software packages (e.g., PROC GLIMMIX in SAS and the `glmmPQL` package in R and S-Plus). These differences depend on whether the scale parameter ϕ is regarded as fixed and known or as an additional parameter to be estimated from the data at hand. Although for many distributions in the exponential family ϕ is fixed and known (e.g., binomial and Poisson, where $\phi = 1$), empirically the variability of the response often exceeds that predicted by these distributions. Thus, when using the PQL method to fit GLMMs in existing software packages, it may be necessary to override the default options concerning whether the scale parameter ϕ is constrained (e.g., $\phi = 1$) or included in the PQL estimation as an additional parameter to be estimated from the data.

15.3 MARGINAL QUASI-LIKELIHOOD

There is a second approximation that leads to a method known as marginal quasi-likelihood (MQL) estimation. Similar to PQL, MQL can be motivated via an approximation of the GLMM. With MQL the approximation is based on a Taylor series expansion of (15.2) around estimates $\widehat{\beta}$ and around $b_i = 0$. For a canonical link function this results in the following approximation to the model given by (15.2):

$$Y_{ij}^{**} \approx X'_{ij}\beta + Z'_{ij}b_i + \epsilon_{ij}^{**}, \tag{15.5}$$

where $Y_{ij}^{**} = v^{-1}(\widehat{\mu}_{ij})(Y_{ij} - \widehat{\mu}_{ij}) + X'_{ij}\widehat{\beta}$ and $\epsilon_{ij}^{**} = v^{-1}(\widehat{\mu}_{ij})\epsilon_{ij}$. Thus we now also have a standard linear mixed effects model for the transformed response, Y_{ij}^{**}, with fixed effects β, random effects b_i, and within-subject errors ϵ_{ij}^{**}. Note that the transformed response in MQL, Y_{ij}^{**}, depends on $\widehat{\mu}_{ij} = g^{-1}(X'_{ij}\widehat{\beta})$, and not on $\widehat{\mu}_{ij}^{b} = g^{-1}(X'_{ij}\widehat{\beta} + Z'_{ij}\widehat{b})$. That is, in contrast to PQL, μ_{ij} depends on the *marginal* linear predictor $(X'_{ij}\beta)$ instead of the *conditional* linear predictor $(X'_{ij}\beta + Z'_{ij}b_i)$. As we will see, this has some important implications for interpretation of the resulting MQL estimates of β.

Recognizing that the "working" response Y_{ij}^{**} follows a linear mixed effects model, with fixed effects β, and with random effects b_i and within-subject errors ϵ_{ij}^{**}, estimation can proceed, similar to PQL, by iterating between the following two steps:

1. Fit a linear mixed effects model to the working response Y_{ij}^{**}, to obtain updated estimates of β and G. The linear mixed effects model is fitted using weights, w_{ij}, that are inversely proportional to the variance of the ϵ_{ij}^{**}; for example, $w_{ij} = \phi^{-1}\, v(\widehat{\mu}_{ij})$, when a canonical link function is adopted.
2. Use the estimates of β obtained in step 1 to update the working response Y_{ij}^{**} (and also update the weights, w_{ij}, used for estimation in step 1).

This two-step algorithm can be iterated until convergence has been achieved. The resulting estimates are known as the MQL estimates.

One crucial aspect of the MQL method is often overlooked. Because the MQL method involves an approximation based on a Taylor series expansion of (15.2) around estimates $\widehat{\beta}$ and around $b_i = 0$, it does not yield an estimate of β in the GLMM:

$$g\{E(Y_{ij}|b_i)\} = X'_{ij}\beta + Z'_{ij}b_i.$$

Instead, because the transformed response in MQL, Y_{ij}^{**}, depends on the *marginal* mean $\widehat{\mu}_{ij} = g^{-1}(X'_{ij}\widehat{\beta})$, and not on the *conditional* mean $\widehat{\mu}_{ij}^{b} = g^{-1}(X'_{ij}\widehat{\beta} + Z'_{ij}\widehat{b})$, the MQL algorithm yields an estimate of the regression parameters β in the following marginal model:

$$g\{E(Y_{ij})\} = X'_{ij}\beta.$$

Thus the MQL method is appropriate only when the goal of the analysis is to make population-averaged, not subject-specific, inferences. Some of the statistical literature on GLMMs, and much of the documentation for commercially available software packages for fitting GLMMs, has been less than transparent about the different targets of inference associated with the PQL and MQL methods. In the next section we highlight how the PQL and MQL methods differ in terms of their respective targets of inference, with the PQL method being appropriate when the goal is to make subject-specific inferences whereas the MQL method is appropriate when the goal is to make population-averaged inferences.

15.4 CAUTIONARY REMARKS ON THE USE OF PQL AND MQL

We note that PQL and MQL both involve approximations that result in a linearization of the GLMM. Consequently both methods can be implemented through repeated use of statistical software for fitting standard linear mixed effects models. This feature of both methods makes them computationally simple to apply. However, the two methods differ in terms of the locus of the linearization of (15.2): in PQL the expansion of (15.2) is around current estimates $\widehat{\beta}$ and $\widehat{b}_i$. In a certain sense, this can be considered a "subject-specific" expansion of (15.2). In contrast, in MQL the expansion of (15.2) is around current estimates $\widehat{\beta}$ and around $b_i = 0$, the mean of the random effects. In a similar way the latter can be considered a "population-averaged" expansion of (15.2). As we discuss in greater detail below, this has ramifications for the interpretation of β when estimated using the PQL and MQL algorithms.

Next we consider the factors that influence the accuracy of the approximation. In the case of PQL, the accuracy of the approximation improves when the "sufficient statistics" for β and b_i have approximate normal distributions. The "sufficient statistics" for β and b_i can be loosely defined in term of those quantities that provide all of the possible information in the sample that is useful for estimating β and b_i. It can be shown that estimation of β relies on the following sufficient statistic: $\sum_{j=1}^{n_i} X_{ij}Y_{ij}$. In contrast, estimation of b_i relies on $\sum_{j=1}^{n_i} Z_{ij}Y_{ij}$. Both of these statistics are weighted averages of the Y_{ij}, where averaging is over the repeated occasions. Thus, in cases where Y_{ij} is quantitative (e.g., counts from a Poisson distribution with mean > 5), we can expect these weighted averages to have approximate normal distributions even when the number of repeated measurements, n_i, is relatively small. Conversely, when Y_{ij} is binomial but based on a small denominator, and especially when Y_{ij} is binary (with binomial denominator of 1), these weighted averages do not have approximate normal distributions unless the number of repeated measurements is very large. This helps explain why PQL often produces very poor estimates of both the fixed effects and the variance components of the random effects when the response variable is binary and there is only a relatively small number of repeated measurements available on each individual. In such settings, PQL can yield seriously biased estimates of effects; in general, the estimates of the fixed effects, but especially the variance components, will be attenuated toward zero. This systematic underestimation of the variance components has been very well documented in the case of binary responses; consequently PQL should be used only in the setting of binomial proportions when the denominator is relatively large and the numerators take on values in the mid-range, for example, binomial proportions where the expected number of successes (or failures) exceeds 5. In principle, as the number of repeated measurements increases, the bias of PQL decreases accordingly.

It is also quite instructive to compare PQL to adaptive Gaussian quadrature. Recall that adaptive Gaussian quadrature can be made as accurate as necessary by increasing the number of quadrature points. Although the relationship between PQL and quadrature methods may not be transparent, it can be shown that a Laplace approximation to the GLMM likelihood corresponds to adaptive Gaussian quadrature with just a single quadrature point. Consequently, given the very close relation between PQL and methods based on a Laplace approximation, it should not be too surprising that this less than optimal choice for the number of quadrature points can lead to seriously biased estimates of effects in GLMMs.

When making statistical inferences about the fixed effects and the variance components, there are important differences between the PQL method and "true" ML estimation based on adaptive Gaussian quadrature. Unlike ML estimation, inferences based on the PQL method cannot rely on the likelihood, e.g., likelihood ratio test statistics are no longer available. Inferences must, however, rely on Wald statistics and confidence intervals. So we must caution the reader that many of the commercially available software packages (e.g., PROC GLIMMIX in SAS) that implement PQL produce output that includes values for the maximized "log likelihood" and various information criteria (e.g., AIC) based on the maximized "log likelihood"; these cannot be used for inference and should be completely ignored. The reported

maximized "log likelihood" applies only to the working response vector Y_i^*; it is not the value of the true "log likelihood" for the fitted GLMM to the response vector Y_i.

In regard to the properties of the MQL method, in much of the statistical literature there is a consensus that MQL produces badly biased estimates of the GLMM parameters unless the variability of the random effects is close to zero. Moreover, unlike PQL, with MQL the bias remains even as the number of repeated measurements increases. What is less clear in the statistical literature, and consequently in documentation for commercially available software packages for fitting GLMMs, is that the MQL method does not yield an estimate of β in the GLMM:

$$g\{E(Y_{ij}|b_i)\} = X'_{ij}\beta + Z'_{ij}b_i.$$

Instead, because MQL is based on a "population-averaged" expansion of (15.2), it produces an estimate of the regression parameters β in the following marginal model:

$$g\{E(Y_{ij})\} = X'_{ij}\beta,$$

that is, a model for the mean response that does not include any random effects but assumes that the marginal mean relationship has the same link function (as for the GLMM). That is, the only similarity in the GLMM and marginal model given above is that the link function, $g(\cdot)$, is adopted by both models; otherwise, the "subject-specific" regression parameters in the GLMM and the "population-averaged" regression parameters in the marginal model are discernibly different (see Chapter 16), especially when the variability of the random effects, b_i, is large. Because the MQL method is based on fitting a marginal model to the longitudinal data, rather than a GLMM, the resulting regression estimates are not expected to be unbiased for the fixed effects in the GLMM. Put simply, when there is a fundamental mismatch between the model of interest and the model that is fit to the data, it cannot be expected that the latter will yield valid estimates of the former. When MQL is used for estimation and inference in the GLMM, the bias with the MQL method arises from the incorrect target of inference (i.e., the *marginal* rather than *conditional* mean) that is modeled. This is because the MQL method estimates the regression parameters in the marginal model,

$$g\{E(Y_{ij})\} = X'_{ij}\beta,$$

and not the fixed effects in the GLMM,

$$g\{E(Y_{ij}|b_i)\} = X'_{ij}\beta + Z'_{ij}b_i.$$

Indeed, it is false to claim MQL to be a legitimate method of estimation for GLMMs! So we caution the reader that many of the leading commercially available software packages for fitting GLMMs include MQL as an optional method of estimation.

Nevertheless, while we cannot recommend the use of MQL for estimation of effects in a GLMM, we note one important potential use of the MQL algorithm: estimation of parameters in a marginal model for the longitudinal response. That is, in settings where it is somewhat more natural or convenient to express the "working

covariance" in a marginal model via the introduction of random effects, MQL can be used to estimate the marginal regression parameters. In this sense, MQL can be regarded as a potentially flexible way to model the "working covariance" in the generalized estimating equations (GEE) approach. We illustrate this use of MQL for estimation of marginal model parameters in Section 15.5.

In summary, the PQL and MQL methods of estimation for GLMMs are based on discernibly different expansions of (15.2) and have distinct targets of inference. Because the PQL method is based on a "subject-specific" expansion of (15.2), it yields estimates of β from the following model for the *conditional* mean:

$$E(Y_{ij}|b_i) = g^{-1}(X'_{ij}\beta + Z'_{ij}b_i),$$

where $g^{-1}(\cdot)$ denotes the inverse link function. The PQL method yields valid estimates of β provided that Y_{ij} is quantitative (e.g., counts from a Poisson distribution with mean greater than 5 or binomial proportions where the expected number of successes exceeds 5). PQL can yield badly biased estimates of β (and the variance components) when the response variable is binary and there are few repeated measurements. In the latter settings, we cannot recommend the use of the PQL method. In contrast to PQL, the MQL method is based on a "population-averaged" expansion of (15.2), and it yields estimates of β from the following model for the *marginal* mean (averaged over the distribution of the random effects):

$$E(Y_{ij}) = g^{-1}(X'_{ij}\beta),$$

with $\text{Var}(Y_{ij}) = v\{E(Y_{ij})\}\,\phi$. As discussed in Chapters 12 and 13, the components of β in this marginal model have "population-averaged" interpretations. Because the subject-specific and population-averaged regression parameters can be discernibly different, we cannot recommend that the MQL method be used for estimation of the fixed effects in a GLMM. Instead, we see the MQL method as having potential use for the estimation of regression parameters in marginal models only. Finally, we note that if a marginal model, rather than a GLMM, is specified for the vector of responses, then the PQL and MQL methods are formally equivalent because they are based on the same expansion. In that setting, both methods can be considered GEE estimators of the marginal model parameters.

To help the reader digest the implications of the choice between the PQL and MQL methods for statistical inference, we present a summary in Table 15.1 that contrasts the true targets of inference (i.e., "subject-specific" or "population-averaged") for the estimated regression parameters versus model specification (GLMM versus marginal model) and method of estimation (PQL versus MQL). From Table 15.1 it is apparent that only the PQL method yields estimates of the fixed effects from a GLMM when a GLMM has been specified for the data at hand. In contrast, when a GLMM has been specified, the MQL method estimates the regression parameters in a marginal model for the vector of responses that has the same link (and variance) function as the conditional distribution of the response in the specified GLMM. Finally, in cases where a marginal model has been specified for the data, the resulting estimates of the marginal regression parameters are valid regardless of the choice of estimation

Table 15.1 Targets of inference for the estimated regression parameters as a function of model specification (GLMM versus marginal model) and method of estimation (PQL versus MQL).

	Model Specification	
Estimation Method	GLMM	Marginal
PQL	Subject-specific	Population-averaged
MQL	Population-averaged[a]	Population-averaged

[a] Yields estimates of regression parameters from marginal model with the same link and variance functions as specified in the GLMM.

method; indeed, for a marginal model specification, the PQL and MQL methods yield identical estimators of the marginal regression parameters.

Many of the issues highlighted in this section can be solidified by a numerical illustration based in part on the example given in Section 14.3 (see Figure 14.1). Assume that N individuals are measured repeatedly at baseline and after treatment with a new drug intended to reduce the risk of disease. Specifically, $n/2$ measurements of the response are obtained at baseline, and $n/2$ measurements post-baseline. To model changes in the response probabilities from baseline to post-baseline, we consider the following logistic regression model, with normally distributed random intercepts:

$$\text{logit}\{E(Y_{ij}|b_i)\} = \beta_1^* + \beta_2^* t_{ij} + b_i,$$

where $t_{ij} = 0$ at baseline and $t_{ij} = 1$ post-baseline ($i = 1, ..., N; j = 1, ..., n$). For a "typical" individual from the population (where a "typical" individual is one with unobserved random effect $b_i = 0$), the log odds of disease at baseline is β_1^*; the log odds of disease following treatment with the new drug is $\beta_1^* + \beta_2^*$.

We can simulate data from this model for different values of N and n for the case where $\beta_1^* = 1.5$, $\beta_2^* = -3.0$, and $\text{Var}(b_i) = 1.0$. Then we estimate the model parameters using PQL and MQL methods. Recall that the PQL method yields estimates of β_1^* and β_2^*, albeit potentially biased estimates. In contrast, the MQL method yields estimates of the regression parameters from the marginal logistic model,

$$\text{logit}\{E(Y_{ij})\} = \beta_1 + \beta_2 t_{ij}.$$

Because of the discreteness of the only covariate ($t_{ij} = 0$ or 1) in the logistic regression model with random intercepts, a logistic regression model also holds for the marginal probabilities (averaged over the distribution of the random effects). However, the regression parameters in the marginal model are attenuated, with $\beta_1 = 1.257$ and $\beta_2 = -2.514$. In Table 15.2 we present the PQL and MQL estimates of the "fixed effects" and variance component $\text{Var}(b_i)$ (for the latter, PQL estimates are presented only) when $N = 5000$ and $n = 4, 8, 16, 32, 64$, and 128. We have

Table 15.2 PQL and MQL estimates, and percent relative bias (in parentheses) when compared to true parameter values, as a function of the number of repeated measurements (n) for simulated data from the logistic regression model with randomly varying intercepts.

	β_1^*	β_1	β_2^*	β_2	$\text{Var}(b_i)$
True Value	1.500	1.257	−3.000	−2.514	1.000
n	PQL	MQL	PQL	MQL	PQL
4	1.310	1.257	−2.619	−2.513	0.647
	(12.7%)	(0.0%)	(12.7%)	(0.0%)	(35.3%)
8	1.370	1.267	−2.752	−2.545	0.780
	(8.7%)	(0.8%)	(8.3%)	(1.2%)	(22.0%)
16	1.415	1.263	−2.844	−2.539	0.864
	(5.7%)	(0.5%)	(5.2%)	(1.0%)	(13.6%)
32	1.443	1.255	−2.902	−2.511	0.947
	(3.8%)	(0.6%)	(3.3%)	(0.1%)	(5.3%)
64	1.470	1.255	−2.939	−2.510	0.962
	(2.0%)	(0.2%)	(2.0%)	(0.1%)	(3.8%)
128	1.490	1.263	−2.972	−2.519	0.971
	(0.7%)	(0.5%)	(0.9%)	(0.2%)	(2.9%)

Note: β_1^* and β_2^* are the regression parameters in the random effects logistic regression model; β_1 and β_2 are the corresponding regression parameters in the marginal logistic regression model (averaged over the distribution of the random effects).

purposely chosen a very large value for N, so that any concerns about sampling variation of the reported estimates can be completely set aside, but have allowed for a broad range of values for n, the number of repeated measures.

Table 15.2 shows how the PQL and MQL methods yield estimates of the regression parameters from GLMM and marginal models, respectively. In the last row of Table 15.2 we can see that the PQL and MQL estimates are very close to the true values of the parameters in the respective random effects and marginal models for these data. This confirms our earlier warning about the different targets of inference for these two different methods of estimation. Table 15.2 also shows that the MQL method yields unbiased estimates of the marginal regression parameters. The relative bias of the MQL estimates of the marginal regression parameters is negligible and less than 1% in most instances. In contrast, the PQL estimates of the fixed effects and variance component are badly biased when the number of repeated measures (n) is relatively small. For example, when $n = 4$, the relative bias of the estimates of the fixed effects is approximately 13%, while the relative bias of the estimate of $\text{Var}(b_i)$ is approximately 35%. The bias of the estimates of the fixed effects remains, and

exceeds 5%, even for $n = 16$. Recall that the marginal probability of success (here denoting disease) at baseline is approximately 0.78. Thus for $n = 16$ the binomial denominator at baseline is $n/2 = 8$, and the expected number of successes and/or failures ($0.78 \times 8 = 6.2$ or $0.22 \times 8 = 1.8$) does not exceed 5. So, in light of our earlier warnings about the use of PQL with binary data, it should not be too surprising that there is discernible bias when $n = 16$ (and the expected number of failures is less than 2). In contrast, when $n = 64$ the binomial denominator at baseline is $n/2 = 32$, and the expected number of failures ($0.22 \times 32 = 7.04$) now exceeds 5; for $n = 64$ the bias is almost negligible (approximately 2% for the estimates of the fixed effects). The pattern of results for the PQL method highlights how the bias diminishes as the number of repeated measurements increases.

15.5 CASE STUDIES

In this section we illustrate the main ideas presented in this chapter by considering generalized linear mixed effects models for analyzing longitudinal data from two studies. The first illustration considers a Poisson regression model, with random effects, for analyzing count data on epileptic seizures from a clinical trial of the anti-epileptic drug, progabide. The second illustration considers a logistic regression model, with random effects, for analyzing data on amenorrhea from a randomized clinical trial of contracepting women. These two data sets were previously analyzed in Section 14.5 using GLMMs fit via adaptive Gaussian quadrature. Here we fit similar models for the conditional means using the PQL method to highlight settings where it should and should not be applied. We also present analyses using the MQL method to highlight how this method yields estimates of regression parameters from a marginal rather than conditional (or subject-specific) model.

Clinical Trial of an Anti-epileptic Drug

The first example involves data from the placebo-controlled clinical trial of 59 epileptic patients, conducted by Leppik et al. (1987). Patients with partial seizures were randomized to either progabide or a placebo, as an adjuvant to the standard anti-epileptic therapy. Prior to treatment the number of epileptic seizures during the preceding 8-week interval was recorded. Counts of epileptic seizures during 2-week intervals before each of four successive post-randomization clinic visits were also recorded. The average rates of seizures (per week) at baseline and in the four post-randomization visits are displayed in Figure 15.1. Note that the mean rates over time in the two treatment groups vary from approximately 3 seizures per week to approximately 5 seizures per week. Thus the expected count of the number of seizures in any 2-week interval is approximately 6 to 10. As we will discuss later, the fact that the mean number of seizures is relatively large has implications for the accuracy of the approximation used in the PQL method.

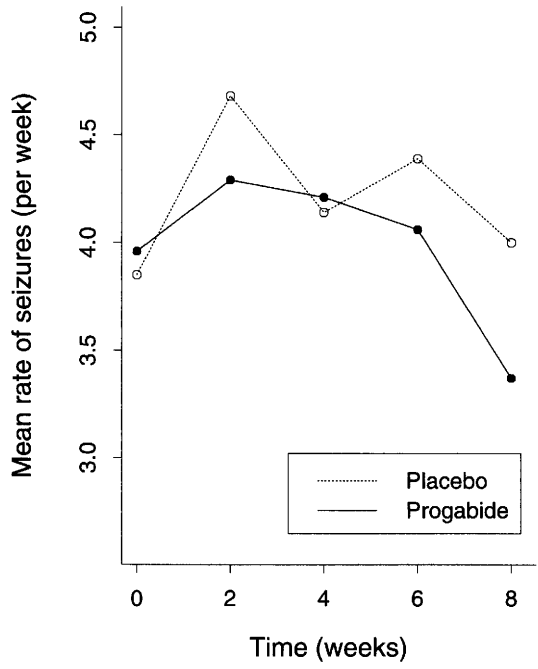

Fig. 15.1 Mean rate of seizures (per week) at baseline, week 2, week 4, week 6, and week 8 in the progabide and placebo groups.

As was noted in Section 14.5, patient 49 is an outlier with extreme counts at all occasions. For illustrative purposes we exclude observations on this patient ($N = 58$) and replicate the analysis displayed in Table 14.7. That is, we consider the following mixed effects log-linear regression model for the subject-specific expected counts (or rates) of seizures:

$$\begin{aligned}\log E(Y_{ij}|b_i) &= \log(\mathrm{T}_{ij}) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\,\mathrm{Time}_{ij} + \beta_3\,\mathrm{Trt}_i \\ &\quad + \beta_4\,\mathrm{Trt}_i \times \mathrm{Time}_{ij},\end{aligned}$$

where Y_{ij} is the number of epileptic seizures for the i^{th} patient in the j^{th} period of observation ($j = 0, \ldots, 4$), and T_{ij} is the length of period j (where $\mathrm{T}_{ij} = 8$ if $j = 0$ and $\mathrm{T}_{ij} = 2$ if $j = 1, 2, 3, 4$). The offset, $\log(\mathrm{T}_{ij})$, is included because the "time at risk" is not the same in the baseline (8 weeks) and four successive follow-up periods (2-week intervals). The variable Trt is an indicator variable for treatment group, with $\mathrm{Trt} = 0$ if an individual was randomized to the placebo group and $\mathrm{Trt} = 1$ if

Table 15.3 PQL estimates and standard errors from mixed effects log-linear regression model for the seizure data, excluding patient 49.

Variable	Estimate	SE	Z
Intercept	1.0869	0.1344	8.09
Time	0.0115	0.1058	0.11
Trt	−0.0074	0.1868	−0.04
Trt × Time	−0.3415	0.1490	−2.29
$g_{11} = \text{Var}(b_{1i})$	0.4579	0.0954	
$g_{22} = \text{Var}(b_{2i})$	0.2196	0.0594	
$g_{12} = \text{Cov}(b_{1i}, b_{2i})$	0.0122	0.0537	

Note: PQL estimation assuming ϕ fixed at 1.

randomized to the progabide group. The binary variable Time denotes the baseline and follow-up periods, with Time = 0 for the baseline period and Time = 1 for the follow-up periods (periods 1–4). Given b_i, it is assumed that the Y_{ij} are independent and have a Poisson distribution, with $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)$. Initially, we assume Poisson variation for $\text{Var}(Y_{ij}|b_i)$ and fix $\phi = 1$. In subsequent analyses, we include ϕ as a parameter to be estimated from the data, allowing for potential overdispersion. Finally, we assume that the random intercepts and slopes, b_i, have a bivariate normal distribution, with zero mean and 2×2 covariance matrix G.

The PQL estimates of the fixed effects and covariance parameters for the log-linear model are displayed in Table 15.3. A test of the null hypothesis, H_0: $\beta_4 = 0$, indicates that there is a significant time × treatment interaction at the 0.05 level. These results suggest that there are differences between the two treatments in terms of subject-specific changes in the expected rates of seizures, with a greater reduction in the expected seizure rate from baseline for a typical patient treated with progabide. Specifically, for a patient receiving a placebo, there is almost no expected change in the rate of seizures (or $1 - e^{0.0115} \approx 0$), while for a patient treated with progabide the expected decrease in seizures is approximately 30% (or $1 - e^{0.0115-0.3415} \approx 0.28$).

It is instructive to compare the PQL estimates of the fixed effects and covariance parameters to the corresponding "true" ML estimates (obtained using 50-point adaptive Gaussian quadrature) in Table 14.7. In general, there is close agreement, indicating that the PQL method provides an adequate approximation in this setting. This is to be expected because, as was noted earlier, the mean of the seizure counts at each occasion is relatively large. Thus for this particular data set the conditions required for the appropriate use of the PQL method are met. We also repeated the PQL analysis but included ϕ as a parameter to be estimated from the data, allowing

Table 15.4 MQL estimates and standard errors from mixed effects log-linear regression model for the seizure data, excluding patient 49.

Variable	Estimate	SE	Z
Intercept	1.3476	0.1574	8.56
Time	0.1118	0.1159	0.96
Trt	−0.1068	0.1937	−0.55
Trt × Time	−0.3024	0.1711	−1.77
$g_{11} = \text{Var}(b_{1i})$	0.5182	0.1043	
$g_{22} = \text{Var}(b_{2i})$	0.3697	0.0834	
$g_{12} = \text{Cov}(b_{1i}, b_{2i})$	−0.0127	0.0660	

Note: MQL estimation assuming ϕ fixed at 1; SE based on empirical variance estimator.

for potential overdispersion. This yielded an estimate of overdispersion, $\widehat{\phi} = 1.94$, suggesting that the conditional variability in the data may be twice as large as that predicted by Poisson variability. However, it should be mentioned that the estimates of the variances of the random effects were noticeably smaller than in Table 15.3, perhaps reflecting that ϕ and the variance component parameters are in competition with each other to explain the variability in the data. The estimate of the fixed effect for the time × treatment interaction, $\widehat{\beta}_4 = -0.303$ (SE = 0.151), is comparable in magnitude to the estimate in Table 15.3 and leads to the same conclusion about the benefits of progabide.

Finally, we re-fit the original model (assuming $\phi = 1$) except now using the MQL method. The results of this analysis are presented in Table 15.4. Note that the estimates of the fixed effects are discernibly different from those displayed in Table 15.3. For example, the estimate of the effect of greatest interest, β_4, is approximately 12% smaller in absolute value. Conversely, the estimate of the intercept is approximately 35% larger. These differences in the estimates of the fixed effects are not a reflection of sampling variability; rather, they reflect the different targets of inference of the PQL and MQL methods. That is, the MQL method does not yield estimates of the fixed effects in the model,

$$\begin{aligned}\log E(Y_{ij}|b_i) &= \log(\mathrm{T}_{ij}) + (\beta_1 + b_{1i}) + (\beta_2 + b_{2i})\,\mathrm{Time}_{ij} + \beta_3\,\mathrm{Trt}_i \\ &\quad + \beta_4\,\mathrm{Trt}_i \times \mathrm{Time}_{ij};\end{aligned}$$

instead, it yields estimates of the regression parameters in the marginal model,

$$\log E(Y_{ij}) = \log(\mathrm{T}_{ij}) + \beta_1 + \beta_2\,\mathrm{Time}_{ij} + \beta_3\,\mathrm{Trt}_i + \beta_4\,\mathrm{Trt}_i \times \mathrm{Time}_{ij},$$

Table 15.5 GEE estimates and standard errors from marginal log-linear regression model for the seizure data, excluding patient 49.

Variable	Estimate	SE	Z
Intercept	1.3312	0.1611	8.26
Time	0.1141	0.0926	1.23
Trt	−0.1021	0.1959	−0.52
Trt × Time	−0.3167	0.1494	−2.12

Note: GEE estimation with unstructured "working" covariance; SE based on empirical variance estimator.

that has the same log link function and variance function, $\text{Var}(Y_{ij}) = E(Y_{ij})$, and with a "working" covariance determined via the introduction of randomly varying intercepts and slopes. To reinforce this point, we used the standard GEE method to fit the marginal model given above to the epilepsy data. The GEE estimates of the regression parameters, obtained under an unstructured "working" covariance, are presented in Table 15.5. As expected, the estimated regression parameters are very close to those obtained using the MQL method. The very small differences in the estimates of the regression parameters are due to the different choices of "working" covariance. In general, with complete data and relatively large sample sizes, we would expect to find no important differences between the estimates from the MQL method and the more standard GEE approach. Indeed, when the MQL method is applied to a GLMM, it should simply be regarded as a GEE method of estimation of the parameters in a marginal model that paradoxically specifies the "working" covariance through the introduction of random effects. Thus, when applying the MQL method to a GLMM, the specification of random effects only has implications for the implicit choice of a "working" covariance for the corresponding marginal model that assumes the same link and variance functions for the marginal means and variances, respectively.

Clinical Trial of Contracepting Women

The next example is from a longitudinal clinical trial of contracepting women reported by Machin et al. (1988). In this trial women received an injection of either 100 mg or 150 mg of depot-medroxyprogesterone acetate (DMPA) on the day of randomization and three additional injections at 90-day intervals. There was a final follow-up visit 90 days after the fourth injection. Throughout the study each woman completed a menstrual diary that recorded any vaginal bleeding pattern disturbances. The diary data were used to determine whether a women experienced amenorrhea, the absence of menstrual bleeding, for a specified number of days.

A total of 1151 women completed the menstrual diaries, and the diary data were used to generate a binary sequence for each woman, according to whether she had

Table 15.6 PQL estimates and standard errors from a mixed effects logistic regression model, with randomly varying intercepts, for the amenorrhea data.

Variable	Estimate	SE	Z
Intercept	−2.5859	0.2298	−11.25
Time	0.7735	0.2178	3.55
Time^2	−0.0267	0.0449	−0.59
Dose × Time	0.3666	0.1416	2.59
Dose × Time^2	−0.0725	0.0385	−1.88
σ_b^2	1.8488	0.1693	

Note: PQL estimation assuming ϕ is fixed at 1.

experienced amenorrhea in the four successive three-month intervals. A feature of this clinical trial is that there was substantial dropout. More than one-third of the women dropped out before the completion of the trial; 17% dropped out after receiving only one injection of DMPA, 13% dropped out after receiving only two injections, and 7% dropped out after receiving three injections. The outcome of interest is a binary response indicating whether a woman experienced amenorrhea in the four successive three-month intervals. The goal of the analyses presented here is to determine subject-specific changes in the risk of amenorrhea over the course of the study (12 months), and the influence of dosage of DMPA on changes in a woman's risk of amenorrhea.

Letting $Y_{ij} = 1$ if the i^{th} woman experienced amenorrhea in the j^{th} injection interval ($j = 1, \ldots, 4$), and $Y_{ij} = 0$ otherwise, we consider the following mixed effects logistic regression model for Y_{ij},

$$\begin{aligned}\text{logit}\{E(Y_{ij}|b_i)\} &= \beta_1 + \beta_2\,\text{Time}_{ij} + \beta_3\,\text{Time}_{ij}^2 + \beta_4\,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5\,\text{Dose}_i \times \text{Time}_{ij}^2 + b_i,\end{aligned}$$

where Time $= 1, 2, 3, 4$ for the four consecutive 90-day injection intervals, and Dose $= 1$ if randomized to 150 mg of DMPA, and Dose $= 0$ otherwise. Due to randomization, we assume that the baseline risk (at Time $= 0$) is the same in both dosage groups and omit a main effect of dose from the model. Given b_i, it is assumed that the Y_{ij} are independent and have a Bernoulli distribution, with $\text{Var}(Y_{ij}|b_i) = E(Y_{ij}|b_i)\{1 - E(Y_{ij}|b_i)\}$, and $\phi = 1$. Finally, we assume that the single random effect b_i has a univariate normal distribution, with zero mean and variance σ_b^2, $b_i \sim N(0, \sigma_b^2)$.

The PQL estimates of the regression parameters for this model are presented in Table 15.6. These results provide evidence that the subject-specific log odds of amenorrhea increases over the 12 months of the trial, and that subject-specific changes in

Table 15.7 MQL estimates and standard errors from a mixed effects logistic regression model, with randomly varying intercepts, for the amenorrhea data.

Variable	Estimate	SE	Z
Intercept	−2.2163	0.1770	−12.52
Time	0.6851	0.1598	4.29
Time^2	−0.0314	0.0322	−0.97
Dose × Time	0.3080	0.1115	2.76
Dose × Time^2	−0.0611	0.0290	−2.11
σ_b^2	1.2608	0.1184	

Note: MQL estimation assuming ϕ is fixed at 1; SE based on empirical variance estimator.

the risk of amenorrhea depend on the dose of DMPA. For example, for a woman assigned to the low dose of DMPA, the log odds of amenorrhea increases approximately linearly, with an increase in the log odds of 0.75 (or $0.7735 - 0.0267$) at 3 months, 1.44 (or $2 \times 0.7735 - 4 \times 0.0267$) at 6 months, 2.08 (or $3 \times 0.7735 - 9 \times 0.0267$) at 9 months, and 2.67 (or $4 \times 0.7735 - 16 \times 0.0267$) at 12 months. These increases in risk correspond to odds ratios of 2.1 (or $e^{0.75}$), 4.2 (or $e^{1.44}$), 8.0 (or $e^{2.08}$), and 14.4 (or $e^{2.67}$) at 3, 6, 9, and 12 months, respectively. Note that the estimates of the fixed effects are discernibly smaller in absolute value than the "true" ML estimates reported in Table 14.2. The difference between the estimates is even more marked for the variance of the random effect. The ML estimate is approximately 5.1 whereas the PQL estimate is approximately 1.8; the former is almost 3 times larger than the latter. This highlights how badly biased the PQL estimates can be when the response is binary and there are relatively few repeated measures. In this setting the PQL method cannot be recommended and inferences based on the PQL estimates are not trustworthy.

Finally, for illustrative purposes only, we also fit the same model using the MQL method. The results of this analysis are presented in Table 15.7. In general, the estimates of the fixed effects and the variance component are attenuated toward zero (in absolute value) relative to the corresponding estimates produced by the PQL method. However, we remind the reader that the PQL method yields estimates, albeit biased estimates, of the regression parameters from the model for the conditional log odds,

$$\begin{aligned}\text{logit}\{E(Y_{ij}|b_i)\} &= \beta_1 + \beta_2\,\text{Time}_{ij} + \beta_3\,\text{Time}_{ij}^2 + \beta_4\,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5\,\text{Dose}_i \times \text{Time}_{ij}^2 + b_i.\end{aligned}$$

In contrast, the MQL method yields perfectly valid estimates of the regression parameters from the following model for the marginal log odds:

$$\begin{aligned}\text{logit}\{E(Y_{ij})\} &= \beta_1 + \beta_2\,\text{Time}_{ij} + \beta_3\,\text{Time}_{ij}^2 + \beta_4\,\text{Dose}_i \times \text{Time}_{ij} \\ &\quad + \beta_5\,\text{Dose}_i \times \text{Time}_{ij}^2.\end{aligned}$$

15.6 COMPUTING: FITTING GENERALIZED LINEAR MIXED MODELS USING PROC GLIMMIX IN SAS

PROC GLIMMIX in SAS is a procedure for fitting GLMMs (among other models) using maximum likelihood (via adaptive quadrature), PQL, or MQL methods; in this section we focus on PQL and MQL methods only. The procedure is very versatile and the syntax is remarkably similar to that used in PROC MIXED. For example, PROC GLIMMIX has a RANDOM statement that is used in a way similar to PROC MIXED for introducing random effects. Although PROC GLIMMIX does not have an explicit REPEATED statement, it does include an option on the RANDOM statement that mirrors how the REPEATED statement is used in PROC MIXED. Recall that the algorithms for both the PQL and MQL methods require repeated use of software for fitting linear mixed effects models, so this close correspondence in syntax should not be too surprising.

Of note, much of the initial output produced by PROC GLIMMIX is standard output from a linear mixed effects model applied to the working response vector, Y_i^* or Y_i^{**}. Inferences should be based only of the reported estimates, standard errors, and Wald statistics. In particular, the reader is cautioned that the reported value of the (pseudo) log-likelihood, and various (pseudo) likelihood-based goodness of fit statistics, in the PQL or MQL output should not be used for making inferences.

Before discussing syntax in any detail, it is important to note that PROC GLIMMIX allows the specification of two distinct types of models and two distinct methods of estimation of the model parameters. Specifically, we can distinguish models in terms of the inclusion or absence of random effects. The former are referred to as "conditional" (i.e., conditional on random effects) or "subject-specific" models; the latter are referred to as "marginal" or "population-averaged" models. Thus PROC GLIMMIX can be used for fitting both GLMMs and marginal models to longitudinal data. PROC GLIMMIX provides two approximate methods of estimation of the model parameters: PQL and MQL. It also provides a method based on the Laplace approximation; as noted earlier in the chapter, a Laplace approximation to the integrand in the likelihood for a GLMM corresponds to the use of adaptive Gaussian quadrature with only a single quadrature point.

If a marginal model has been specified, both PQL and MQL can be considered GEE approaches, and the two methods will yield identical estimates of the marginal regression parameters and the "working" covariance. So, in cases where a marginal model has been specified, there is no distinction between the PQL and MQL methods. However, in cases where a GLMM has been specified, there are important differences

between PQL and MQL. As discussed in earlier sections of this chapter, only the PQL method yields estimates, albeit sometimes biased estimates, of the fixed effects in a GLMM. The MQL method emphatically does not yield estimates of the fixed effects in a GLMM; instead, it provides estimates of the regression parameters in a marginal model that has the same assumed link and variance functions (e.g., logit link function and Bernoulli variance for the analysis of binary responses) for the marginal means and variances, respectively.

In distinguishing PQL and MQL we note also that PROC GLIMMIX provides two small variations of these two methods of estimation: "restricted maximum" and "maximum" PQL and MQL estimation. As noted in Sections 15.2 and 15.3, the PQL and MQL algorithms solve the linear mixed effects model likelihood equations for a "working" vector of responses, denoted Y_i^* or Y_i^{**}, respectively; thus, as with standard linear mixed effects models, estimation of the covariance of the random effects (or the marginal covariance) can be based on either the REML or ML equations for these parameters. "Restricted maximum" PQL and MQL refer to the algorithms that use the REML equations for estimation of the covariance parameters; "maximum" PQL and MQL refer to the algorithms that use the ML equations. In PROC GLIMMIX, the former are denoted RSPL and RMPL, the latter are denoted MSPL and MMPL. These abbreviations are not very intuitive but can be deciphered as follows. The last two letters, PL, denote that the PQL and MQL methods of estimation are based on a so-called pseudo-likelihood for a linearization (or approximation) to the model. The first letter distinguishes between "restricted maximum" (R) and "maximum" (M) PQL and MQL estimation of the covariance parameters. Finally, the second letter distinguishes between the PQL (S) and MQL (M) methods, recognizing that the former method is based on a "subject-specific" (S) expansion of the model, while the latter is based on a "marginal" (M) or "population-averaged" expansion.

The possible choices of methods of estimation, and their implications for inference (e.g., "subject-specific" or "population-averaged"), are summarized in Table 15.8. For example, the fixed effects (and variance components) in a GLMM can be estimated using either restricted or maximum PQL methods through the use of the METHOD=RSPL or METHOD=MSPL options, respectively. In the absence of random effects in the model, there is no distinction between PQL and MQL methods for estimating the parameters in a marginal model. Thus, when PROC GLIMMIX is used for fitting marginal models, the choice of either METHOD=RSPL or METHOD=RMPL yields identical estimates and standard errors; similarly the choice of either METHOD=MSPL or METHOD=MMPL produces identical results for fitting marginal models.

One appealing feature of PROC GLIMMIX is that it can fit a broad and flexible class of models. To do so, PROC GLIMMIX makes a distinction between two sources of variation and covariation in the model for the data: (1) variation due to random effects, b_i, and (2) "residual" (co)variation. To distinguish these two sources of (co)variation, PROC GLIMMIX refers to (1) as "G-side" effects and (2) as "R-side" effects. These two non-standard terms are derived from notation for the covariance matrices for the random effects (denoted G) and the "residual errors" (denoted R) in PROC MIXED for linear mixed effects models. PROC GLIMMIX is quite versatile in allowing models

Table 15.8 Targets of inference for the estimated regression parameters as a function of model specification (GLMM versus marginal model) and method of estimation (PQL versus MQL) using the METHOD=<option> in PROC GLIMMIX in SAS.

	Model Specification	
Estimation Method	GLMM	Marginal
PQL	Subject-specific	Population-averaged
Restricted Maximum PQL	(RSPL)	(RSPL or RMPL)
Maximum PQL	(MSPL)	(MSPL or MMPL)
MQL	Population-averaged[a]	Population-averaged
Restricted Maximum MQL	(RMPL)	(RMPL or RSPL)
Maximum MQL	(MMPL)	(MMPL or MSPL)

[a] Yields estimates of regression parameters from a marginal model with the same link and variance functions as specified in the GLMM.

to be fit with various combinations of "G-side" and "R-side" effects. For example, in a standard marginal model, there are no random effects b_i; consequently "G-side" effects are completely absent, and the marginal covariance is ordinarily specified in terms of "R-side" effects only. In a standard GLMM, the introduction of random effects is handled by including "G-side" effects, and the "conditional independence" assumption is handled by assuming a simple structure for the "R-side" effects (i.e., uncorrelated residual errors, the default option for the "R-side" effects). In principle, it is possible to fit models that relax the "conditional independence" assumption by allowing for a more general structure for the "R-side" effects (e.g., autoregressive residual errors); however, we caution that, as with linear mixed effects models, there can be subtle issues of model identification when a more general structure for the "R-side" effects is assumed because it may not be possible to estimate both the "G-side" and the more general "R-side" effects from the data at hand.

Because the "G-side" and "R-side" terminology has been borrowed from PROC MIXED, it is useful to compare how PROC MIXED and PROC GLIMMIX specify these effects. In PROC MIXED, the RANDOM statement is used to specify the random effects b_i (the "G-side" effects). In PROC MIXED, multiple RANDOM statements are allowed. Similarly, in PROC GLIMMIX, the RANDOM statement is used to specify the random effects b_i ("G-side" effects) in a GLMM; PROC GLIMMIX also allows the use of multiple RANDOM statements. In PROC MIXED, the REPEATED statement is used to specify assumptions about the "residual errors" ("R-side" effects) in models with or without random effects. This is where the syntax for PROC GLIMMIX departs from PROC MIXED. PROC GLIMMIX does not have a REPEATED statement. Instead, it includes a refinement to the RANDOM

Table 15.9 Illustrative commands for a mixed effects logistic regression, with randomly varying intercepts, fitted using PQL in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=RSPL;
    CLASS id group;
    MODEL y=group time group*time / DIST=BINOMIAL LINK=LOGIT S;
    RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
```

Table 15.10 Illustrative commands for a mixed effects log-linear regression, with randomly varying intercepts and slopes, fitted using PQL in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=RSPL;
    CLASS id group;
    MODEL y=group time group*time / DIST=POISSON LINK=LOG S;
    RANDOM INTERCEPT time / SUBJECT=id TYPE=UN;
```

statement that makes it completely equivalent to a REPEATED statement, via the use of the RESIDUAL option or through the inclusion of a _RESIDUAL_ effect (these options will be explained in detail below). So, unlike in PROC MIXED where both a RANDOM and REPEATED statement can be used in specifying a linear mixed effects model, in PROC GLIMMIX assumptions about the random effects and the "residual errors" are made using two RANDOM statements, with the second being a variant of the RANDOM statement that includes an option or reserved keyword that signifies that it is referring to the "R-side" effects rather than the "G-side" effects.

In many other respects, though, PROC MIXED and PROC GLIMMIX have a lot of similarity in terms of command syntax. For example, to fit a logistic regression model with randomly varying intercepts to longitudinal data from two groups using "restricted" PQL, we can use the illustrative SAS commands given in Table 15.9. Similarly, to fit a mixed effects log-linear regression, with randomly varying intercepts and slopes (also via "restricted" PQL), we can use the illustrative SAS commands given in Table 15.10.

To fit a marginal logistic regression model, rather than a mixed effects model, to longitudinal data from two groups, we can use the illustrative SAS commands given in Table 15.11; note that for a marginal model specification, the options METHOD=RSPL and METHOD=RMPL yield identical results. This is the most direct and transparent way of fitting a marginal model, since the absence of any random effects is denoted by the use of the RESIDUAL option on the single RAN-

Table 15.11 Illustrative commands for a marginal logistic regression model, with exchangeable working correlation, fitted using PQL/MQL in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=RSPL;
    CLASS id group occasion;
    MODEL y=group time group*time / DIST=BINOMIAL LINK=LOGIT S;
    RANDOM occasion / SUBJECT=id TYPE=CS RESIDUAL;
```

Table 15.12 Illustrative commands for a marginal logistic regression model, with working correlation specified by introducing randomly varying intercepts, fitted using MQL in PROC GLIMMIX in SAS.

```
PROC GLIMMIX METHOD=RMPL;
    CLASS id group;
    MODEL y=group time group*time / DIST=BINOMIAL LINK=LOGIT S;
    RANDOM INTERCEPT / SUBJECT=id TYPE=UN;
```

DOM statement. However, recognizing that the MQL method is simply another GEE method for estimating marginal model parameters, we can also fit the marginal logistic regression model using the illustrative SAS commands given in Table 15.12. Note that although the model is specified to be a GLMM with randomly varying intercepts, MQL estimates are requested by using the option METHOD=RMPL instead of METHOD=RSPL. This implies that the resulting MQL estimates of the regression parameters refer not to the logistic regression model with randomly varying intercepts but to the marginal logistic regression model. Although this approach to estimation of marginal model parameters is somewhat less transparent, because it invokes a GLMM, it may have potential applications in settings where the data are highly unbalanced over time. With inherently unbalanced data, it is somewhat more appealing to specify a "working" covariance via the inclusion of randomly varying intercepts and slopes; that is, the "working" covariance is expressed as an explicit function of the times of measurement when times of measurement are included in the covariates for the random effects (Z_i). Next we present a brief description of each of the command statements used in Tables 15.9 through 15.12.

PROC GLIMMIX <options>;

This statement calls the procedure GLIMMIX in SAS. It includes an option for specifying the method of estimation, using METHOD=<options>. The

default is METHOD=RSPL; three other options include METHOD=MSPL, METHOD=RMPL, and METHOD=MMPL (see Table 15.8). PROC GLIMMIX also includes an implementation of adaptive Gaussian quadrature using the METHOD=QUAD option. There is also an EMPIRICAL=CLASSICAL option that requests that standard errors and test statistics for the fixed effects be based on the classical empirical or "sandwich" estimator; this option is particularly useful when PROC GLIMMIX is used to fit marginal models. PROC GLIMMIX also includes options for various bias-corrected versions of the empirical variance estimator.

CLASS variables;

The CLASS statement is used to define all variables that are to be treated as categorical or factors. By default, this statement will create indicator variables for each factor using a reference group coding, with the last level (where "last" here refers to the level with the largest alphanumeric value) regarded as the reference group. Different sort orders for the CLASS variables can be requested by the ORDER=<option> on the PROC GLIMMIX statement.

MODEL response = <effects> / <options>;

MODEL events/trials = <effects> / <options>;

The MODEL statement specifies the response variable and the covariate effects. The second form of the MODEL statement, with the `events/trials` syntax, allows the response to be in the form of a ratio of two variables (e.g., counts of the number of successes and the number of trials) and is used for binomial response data. The linear predictor can include both discrete (defined in the CLASS statement) and quantitative (excluded from the CLASS statement) covariates. By default, PROC GLIMMIX includes a column of 1's for the intercept in the model. The SOLUTION or S option is used to produce estimates of the fixed (or covariate) effects.

The option DIST=*keyword* specifies the conditional distribution of the response given the random effects in a GLMM. Alternatively, if PROC GLIMMIX is used to fit a marginal model, the DIST=*keyword* has a somewhat different role. In fitting a marginal model using the GEE approach, the DIST=*keyword* does not specify the distribution for the vector of correlated responses; instead, it specifies the default canonical link function and variance function that happen to be associated with particular exponential family distributions. For example, for a marginal model the option DIST=POISSON does not specify that the response vector (or even its separate components) has a Poisson distribution; instead, it specifies that the mean of the response vector is related to the covariates via a log link function (the canonical link for the Poisson distribution) and the mean and variance of the responses are related by $\text{Var}(Y) = E(Y) = \mu$ (i.e., the variance function is $v(\mu) = \mu$).

Note that PROC GLIMMIX provides a wide choice of options for the inclusion of a dispersion parameter, ϕ. However, unlike PROC GENMOD, the option for the inclusion of a dispersion parameter does not appear on the MODEL

statement; instead, a dispersion parameter can be included by using the "G-side" variant of the RANDOM statement (see below).

The LINK=*keyword* specifies the choice of built-in link function relating the mean response to the linear predictor. If the LINK=*keyword* is omitted, the default link function is the canonical link function associated with the particular exponential family distribution specified on DIST=*keyword*.

A final option often required when modeling count data is an offset. The OFFSET=*variable* specifies a variable to be used as an offset. For example, in modeling count data the rate is often of more direct interest and the denominator for the counts or "population at risk" (or more specifically, the log of the denominator) can be included as an offset. Note that this variable cannot be a CLASS variable and it should not be included as one of the covariates listed on the MODEL statement.

RANDOM <random-effects> / SUBJECT=subject-effect <options>;

The RANDOM statement is used in two different ways in PROC GLIMMIX: it is used to define the random effects, b_i ("G-side" effects), and also used to specify the "residual errors" ("R-side" effects).

In a GLMM, the RANDOM statement is used to define the covariates in the design matrix, Z_i, for the random effects, b_i. Ordinarily these will be a subset of the covariates included on the MODEL statement. While the MODEL statement is used to define the design matrix for the fixed effects and the RANDOM statement is used to define the design matrix for the random effects, note that an intercept is included by default in the former but not the latter. That is, unlike the MODEL statement, PROC GLIMMIX does not include an intercept in the RANDOM statement by default. However, you can specify INTERCEPT (or INT) as a random effect on the RANDOM statement. The RANDOM statement is also used to specify the structure of the covariance matrix for the random effects, G. The structure of G is specified using the TYPE=option. The random effects can be assumed to be correlated (TYPE=UN) or uncorrelated (TYPE=VC); ordinarily covariance pattern models are not used to account for the covariance among the random effects.

As was noted earlier, unlike PROC MIXED, the GLIMMIX procedure does not have a REPEATED statement for specifying the "residual errors" ("R-side" effects). Instead, it uses the RANDOM statement with either the RESIDUAL option or the reserved _RESIDUAL_ keyword in place of the random effects. The use of the RESIDUAL option on the RANDOM statement indicates that the statement refers to the "residual errors" or "R-side" effects. In most respects, the RANDOM statement with the RESIDUAL option is equivalent to the REPEATED statement in PROC MIXED. For example, when using the RANDOM statement with the RESIDUAL option, it is possible to include a variable denoting the "repeated effect." In the context of longitudinal data, the "repeated effect" is often used to identify the order of the repeated measurements within subjects (see Table 15.11):

RANDOM <repeated effect> / SUBJECT=subject-effect RESIDUAL;

There is also a _RESIDUAL_ keyword that can be used on the RANDOM statement (in place of the <random-effects>):

RANDOM _RESIDUAL_ / SUBJECT=subject-effect;

To recap, the statement:

REPEATED / SUBJECT=id TYPE=CS;

in PROC MIXED is equivalent to the following statement in PROC GLIMMIX:

RANDOM _RESIDUAL_ / SUBJECT=id TYPE=CS;

Similarly, if it is necessary to include a "repeated effect" that identifies the measurement occasions, then the following statements in PROC MIXED:

CLASS id occasion;
REPEATED occasion / SUBJECT=id TYPE=UN;

are equivalent to the following statements in PROC GLIMMIX:

CLASS id occasion;
RANDOM occasion / SUBJECT=id TYPE=UN RESIDUAL;

Finally, simply adding the _RESIDUAL_ keyword to the RANDOM statement:

RANDOM _RESIDUAL_;

specifies a single variance parameter for the "residual errors." That is, the scale parameter ϕ is no longer regarded as fixed; instead it is estimated from the data (e.g., to allow for overdispersion relative to binomial or Poisson variation).

15.7 BASIS OF PQL AND MQL APPROXIMATIONS*

In Section 15.4 it was noted that the PQL and MQL methods are based on two different approximations to the GLMM. In both cases these approximations yield a standard linear mixed effects model for a "working" or transformed response, denoted Y_{ij}^{*} and Y_{ij}^{**} for PQL and MQL, respectively. This section provides a brief description of the basis for the approximations used by the PQL and MQL methods. This section is somewhat technical and can be skipped without loss of continuity.

Recall that the model for the conditional mean of a GLMM,

$$g\{E(Y_{ij}|b_i)\} = \eta_{ij}^b = X_{ij}'\beta + Z_{ij}'b_i, \tag{15.6}$$

can also be expressed as

$$Y_{ij} = g^{-1}(X_{ij}'\beta + Z_{ij}'b_i) + \epsilon_{ij} = \mu_{ij}^b + \epsilon_{ij}, \tag{15.7}$$

where $g^{-1}(\cdot)$ is the inverse link function, μ_{ij}^b denotes the *conditional* mean of Y_{ij}, given b_i, and ϵ_{ij} denotes a mean zero random error. Also we assume $\text{Var}(Y_{ij}|b_i) = \text{Var}(\epsilon_{ij}) = v\{E(Y_{ij}|b_i)\}\,\phi$, where $v(\cdot)$ is a known variance function.

The basis of the PQL method is an approximation to the model given by (15.7) so that the random effects b_i and the within-subject errors enter into the approximate model in an additive, linear fashion. This type of approximation can be achieved by using what is known as a first-order Taylor series expansion of (15.7) around current estimates $\widehat{\beta}$ and $\widehat{b}_i$. Taylor series expansions require an understanding of calculus, so we omit many of the technical details here. It will suffice to mention that a first-order Taylor series expansion can be used to express a non-linear function of β and b_i (e.g., $g^{-1}(X_{ij}'\beta + Z_{ij}'b_i)$) as a sum or linear combination of β and b_i. Specifically, this yields the following approximation:

$$\begin{aligned} Y_{ij} &= g^{-1}(X_{ij}'\beta + Z_{ij}'b_i) + \epsilon_{ij} \\ &\approx g^{-1}(X_{ij}'\widehat{\beta} + Z_{ij}'\widehat{b}_i) + \delta(\widehat{\mu}_{ij}^b)\{X_{ij}'(\beta - \widehat{\beta}) + Z_{ij}'(b_i - \widehat{b}_i)\} + \epsilon_{ij} \\ &= \widehat{\mu}_{ij}^b + \delta(\widehat{\mu}_{ij}^b)\{X_{ij}'(\beta - \widehat{\beta}) + Z_{ij}'(b_i - \widehat{b}_i)\} + \epsilon_{ij}, \end{aligned}$$

where $\delta(\mu_{ij}^b)$ denotes the derivative of μ_{ij}^b with respect to the linear predictor, $\eta_{ij}^b = X_{ij}'\beta + Z_{ij}'b_i$ (in calculus, the "derivative" of a function describes its rate of change, or how quickly that function changes, with respect to some variable). When a canonical link function has been chosen (e.g., logit link function for Bernoulli or log link function for Poisson), $\delta(\mu_{ij}^b) = v(\mu_{ij}^b)$, the variance function. The final step is to re-arrange the terms given above so that all of the unknown quantities appear on the right-hand side:

$$\delta^{-1}(\widehat{\mu}_{ij}^b)(Y_{ij} - \widehat{\mu}_{ij}^b) + X_{ij}'\widehat{\beta} + Z_{ij}'\widehat{b}_i \approx X_{ij}'\beta + Z_{ij}'b_i + \delta^{-1}(\widehat{\mu}_{ij}^b)\epsilon_{ij},$$

where $\delta^{-1}(\widehat{\mu}_{ij}^b) = 1/\delta(\widehat{\mu}_{ij}^b)$. Closer inspection of the right-hand side of the equation above reveals that it conforms to a standard linear mixed effects model, with fixed effects β, random effects b_i, and within-subject errors $\delta^{-1}(\widehat{\mu}_{ij}^b)\epsilon_{ij}$. If we denote the left-hand side of the equation by

$$Y_{ij}^* = \delta^{-1}(\widehat{\mu}_{ij}^b)(Y_{ij} - \widehat{\mu}_{ij}^b) + X_{ij}'\widehat{\beta} + Z_{ij}'\widehat{b}_i,$$

and let $\epsilon_{ij}^* = \delta^{-1}(\widehat{\mu}_{ij}^b)\epsilon_{ij}$, the equation can be expressed as

$$Y_{ij}^* \approx X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij}^*,$$

where the within-subject errors, ϵ_{ij}^*, have mean of zero and variance equal to $\phi[\delta^{-1}(\widehat{\mu}_{ij}^b)]^2 v(\widehat{\mu}_{ij}^b)$. When expressed in this way, the relation to linear mixed effects models is far more transparent. That is, we now have a standard linear mixed effects model for the "working" response (or so-called pseudo-data), Y_{ij}^*, with fixed effects β, and with the random effects, $b_i \sim N(0, G)$, and the within-subject errors, ϵ_{ij}^*, entering into the model in an additive, linear fashion. As a result estimation can proceed by iteratively fitting a linear mixed effects model to the updated "working" response Y_{ij}^* using the 2-step algorithm outlined in Section 15.2.

Next we illustrate the form of the "working" response Y_{ij}^* for a mixed effects logistic regression model. Consider the following logistic regression model with randomly varying intercepts (or subject effects):

$$\log\left\{\frac{\Pr(Y_{ij}=1|b_i)}{\Pr(Y_{ij}=0|b_i)}\right\} = \eta_{ij}^b = X_{ij}'\beta + b_i,$$

where the single random effect b_i is assumed to have a zero mean univariate normal distribution. The model for the conditional mean can be expressed in the following equivalent way:

$$\Pr(Y_{ij}=1|b_i) = E(Y_{ij}|b_i) = \mu_{ij}^b = \frac{\exp(X_{ij}'\beta + b_i)}{1+\exp(X_{ij}'\beta + b_i)}.$$

Conditional on b_i, the Y_{ij} are assumed to be independent and to have a Bernoulli distribution, with

$$\text{Var}(Y_{ij}|b_i) = v\{E(Y_{ij}|b_i)\} = E(Y_{ij}|b_i)\left\{1 - E(Y_{ij}|b_i)\right\},$$

that is, $v(\mu_{ij}^b) = \mu_{ij}^b(1-\mu_{ij}^b)$. Finally, because the canonical link function for the Bernoulli distribution has been adopted,

$$\delta(\mu_{ij}^b) = v(\mu_{ij}^b) = \mu_{ij}^b(1-\mu_{ij}^b).$$

For current estimates of β and b_i, the "working" response Y_{ij}^* is then given by the expression

$$\begin{aligned} Y_{ij}^* &= \delta^{-1}(\widehat{\mu}_{ij}^b)(Y_{ij} - \widehat{\mu}_{ij}^b) + X_{ij}'\widehat{\beta} + \widehat{b}_i \\ &= \frac{Y_{ij} - \widehat{\mu}_{ij}^b}{\widehat{\mu}_{ij}^b(1-\widehat{\mu}_{ij}^b)} + X_{ij}'\widehat{\beta} + \widehat{b}_i, \end{aligned}$$

where

$$\widehat{\mu}_{ij}^b = \frac{\exp(X_{ij}'\widehat{\beta} + \widehat{b}_i)}{1+\exp(X_{ij}'\widehat{\beta} + \widehat{b}_i)}.$$

The steps for deriving the MQL method are very similar except that the approximation is based on a first-order Taylor series expansion of (15.7) around current estimates $\widehat{\beta}$ and $b_i = 0$ (the mean of the random effects). The expansion yields the following approximation:

$$\begin{aligned} Y_{ij} &= g^{-1}(X'_{ij}\beta + Z'_{ij}b_i) + \epsilon_{ij} \\ &\approx g^{-1}(X'_{ij}\widehat{\beta}) + \delta(\widehat{\mu}_{ij})\{X'_{ij}(\beta - \widehat{\beta}) + Z'_{ij}b_i\} + \epsilon_{ij} \\ &= \widehat{\mu}_{ij} + \delta(\widehat{\mu}_{ij})\{X'_{ij}(\beta - \widehat{\beta}) + Z'_{ij}b_i\} + \epsilon_{ij}, \end{aligned}$$

where $\delta(\mu_{ij})$ denotes the derivative of the *marginal* mean, $\mu_{ij} = g^{-1}(X'_{ij}\beta)$, with respect to the *marginal* model linear predictor, $\eta_{ij} = X'_{ij}\beta$. When a canonical link function has been chosen (e.g., logit link function for Bernoulli or log link function for Poisson), $\delta(\mu_{ij}) = v(\widehat{\mu}_{ij})$, the variance function (applied to the marginal mean, μ_{ij}). Note that unlike PQL, μ_{ij} depends on the *marginal* linear predictor $X'_{ij}\beta$ instead of the *conditional* linear predictor $X'_{ij}\beta + Z'_{ij}b_i$. As we have already discussed in earlier sections, this has some important implications for interpretation of the resulting MQL estimates of β.

The final step is to re-arrange the terms given above so that all of the unknown quantities appear on the right-hand side:

$$\delta^{-1}(\widehat{\mu}_{ij})(Y_{ij} - \widehat{\mu}_{ij}) + X'_{ij}\widehat{\beta} \approx X'_{ij}\beta + Z'_{ij}b_i + \delta^{-1}(\widehat{\mu}_{ij})\epsilon_{ij},$$

where $\delta^{-1}(\widehat{\mu}_{ij}) = 1/\delta(\widehat{\mu}_{ij})$. Closer inspection of the right-hand side of the equation above reveals that it conforms to a standard linear mixed effects model, with fixed effects β, random effects b_i, and within-subject errors $\delta^{-1}(\widehat{\mu}_{ij})\epsilon_{ij}$. If we denote the left-hand side of the equation by

$$Y^{**}_{ij} = \delta^{-1}(\widehat{\mu}_{ij})(Y_{ij} - \widehat{\mu}_{ij}) + X'_{ij}\widehat{\beta},$$

and let $\epsilon^{**}_{ij} = \delta^{-1}(\widehat{\mu}_{ij})\epsilon_{ij}$, the equation can be expressed as

$$Y^{**}_{ij} \approx X'_{ij}\beta + Z'_{ij}b_i + \epsilon^{**}_{ij},$$

where the within-subject errors, ϵ^{**}_{ij}, have mean of zero and variance equal to $\phi[\delta^{-1}(\widehat{\mu}_{ij})]^2 v(\widehat{\mu}_{ij})$. That is, we now have a standard linear mixed effects model for the "working" response, Y^{**}_{ij}, with fixed effects β, and with the random effects, $b_i \sim N(0, G)$, and the within-subject errors ϵ^{**}_{ij} entering into the model in an additive, linear fashion. As with the PQL method, estimation can proceed by iteratively fitting a linear mixed effects model to the updated "working" response Y^{**}_{ij}.

Finally, although the PQL and MQL methods both use GLS estimators,

$$\widehat{\beta}^{*} = \left\{ \sum_{i=1}^{N} \left(X'_i V_i^{*-1} X_i \right) \right\}^{-1} \sum_{i=1}^{N} \left(X'_i V_i^{*-1} Y_i^{*} \right)$$

and

$$\widehat{\beta}^{**} = \left\{ \sum_{i=1}^{N} \left(X'_i V_i^{**-1} X_i \right) \right\}^{-1} \sum_{i=1}^{N} \left(X'_i V_i^{**-1} Y_i^{**} \right),$$

respectively, where Y_i^* and Y_i^{**} denote the $n_i \times 1$ vectors of working responses (V_i^* and V_i^{**} are the marginal covariance of Y_i^* and Y_i^{**} respectively), the former can be shown to be based on a weighted average of *conditional* (i.e., conditional on the random effects, b_i) "residuals," $(Y_{ij} - \mu_{ij}^b)$, while the latter is based on a weighted average of *marginal* "residuals," $(Y_{ij} - \mu_{ij})$. Consequently the GLS estimator of β based on Y_i^* (PQL estimator) yields estimates of the fixed effects in the model for the conditional mean,

$$g\{E(Y_{ij}|b_i)\} = \eta_{ij}^b = X_{ij}'\beta^* + Z_{ij}'b_i.$$

In contrast, the GLS estimator of β based on Y_i^{**} (MQL estimator) yields estimates of the regression parameters in the following model for the marginal mean,

$$g\{E(Y_{ij})\} = \eta_{ij} = X_{ij}'\beta^{**}.$$

As was discussed in earlier sections of the book, $\beta^* \neq \beta^{**}$ when a non-linear link function, $g(\cdot)$, is adopted (e.g., logit link function). Thus the PQL method should be used when the goal of the analysis is to make subject-specific inferences for the parameters in the GLMM, whereas the MQL method should be used when the goal is to make population-averaged inferences for the regression parameters in a marginal model that assumes the same link function $g(\cdot)$.

15.8 FURTHER READING

Breslow (2005) presents a concise, and remarkably clear, review of the statistical literature on approximate methods for estimation and inference for generalized linear mixed effects models.

Bibliographic Notes

The theoretical foundations for approximate methods for generalized linear mixed effects models can be found in Stiratelli, Laird, and Ware (1984), Schall (1991), Breslow and Clayton (1993), and Wolfinger (1993). For the special case of the logit-normal model, Stiratelli, Laird, and Ware (1984) proposed an approximate method of estimation, based on empirical Bayes ideas, that circumvented the need for numerical integration. Specifically, they avoided the need for numerical integration by approximating the integrand with a simple expansion whose integral has a closed form. The paper by Stiratelli et al. (1984) provided the impetus for the development of a general approach for fitting generalized linear mixed models known as penalized quasi-likelihood (PQL). Schall (1991), Breslow and Clayton (1993), and Wolfinger (1993), motivated PQL as a Laplace approximation to the marginal likelihood for generalized linear mixed models and highlighted the generality of the PQL method.

Problems

15.1 In a randomized, double-blind, parallel-group, multicenter study comparing two oral anti-fungal treatments (200 mg/day Itraconazole and 250 mg/day Terbinafine) for toenail infection (De Backer et al., 1998; also see Lesaffre and Spiessons, 2001), patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail-bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome variable is binary ("none or mild" versus "moderate or severe"). The binary outcome was evaluated on 294 patients comprising a total of 1908 measurements. The main objective of the analyses is to compare the effects of the two oral anti-fungal treatments (Itraconazole and Terbinafine) on changes in the probability of the binary onycholysis outcome over the duration of the study.

The raw data are stored in an external file: `toenail.dat`

Each row of the data set contains the following five variables:

ID Y Treatment Month Visit

Note: The binary onycholysis outcome variable Y is coded 0 = none or mild, 1 = moderate or severe. The categorical variable Treatment is coded 1 = Terbinafine, 0 = Itraconazole. The variable Month denotes the exact timing of measurements in months. The variable Visit denotes the visit number (visit numbers 1–7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks).

15.1.1 Consider a generalized linear mixed model, with randomly varying intercepts, for the patient-specific log odds of moderate or severe onycholysis. Using penalized quasi-likelihood, fit a model with linear trends for the log odds over time, with common intercept for the two treatment groups, but different slopes:

$$\text{logit}\{E(Y_{ij}|b_i)\} = (\beta_1 + b_i) + \beta_2\,\text{Month}_{ij} + \beta_3\,\text{Treatment}_i \times \text{Month}_{ij},$$

where, given b_i, Y_{ij} is assumed to have a Bernoulli distribution. Assume that $b_i \sim N(0, \sigma_b^2)$.

15.1.2 What is the interpretation of the estimate of β_2?

15.1.3 What is the interpretation of the estimate of β_3?

15.1.4 From the results of the analysis for Problem 15.1.1, what conclusions do you draw about the effect of treatment on changes in the log odds of moderate or severe onycholysis over time? Provide results that support your conclusions.

15.1.5 Repeat the analysis from Problem 15.1.1, fitting the model using maximum likelihood (ML) with 30 point numerical quadrature.

15.1.6 Compare and contrast the estimates of β and σ_b^2 from fitting the model using ML and PQL. Can you explain why they might differ?

15.1.7 Repeat the analysis from Problem 15.1.1, fitting the model using marginal quasi-likelihood (MQL) instead of penalized quasi-likelihood (PQL).

15.1.8 Compare and contrast the estimates of β, especially β_3, from fitting the model using MQL and PQL. Can you explain why the estimates might differ? Can you provide results from an additional analysis of these data that might support your explanation for the difference between the MQL and PQL estimates?