9. The generalized linear mixed model

Sonja Greven

Summer Term 2015

Overview Chapter 9 - The generalized linear mixed model

9.1 Introduction to GLMMs

9.2 Interpretation of the parameters

9.3 Estimation and inference

9.4 Implementation and examples

Generalized linear mixed models (GLMMs)

- **Conditional** on the random effects $\boldsymbol{b}_i$, $Y_{ij}, j = 1, \ldots, n_i$, are **independent** and follow a distribution from the exponential family:

$$f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp\{\phi^{-1}[y_{ij}\theta_{ij} - \psi(\theta_{ij})] + c(y_{ij}, \phi)\}.$$

- The conditional mean $\mu_{ij} = \mathsf{E}(Y_{ij}|\boldsymbol{b}_i)$ is given by

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i,$$

  where $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ are p- and q-vectors of known covariates and g is a known link function. With the natural link, $\theta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i$.

- $\boldsymbol{b}_i$ i.i.d. multivariate normal: $\mathbf{b}_i \overset{iid}{\sim} \mathcal{N}_q(\mathbf{0}, \mathbf{D})$.

Generalized linear mixed models (GLMMs)

$$f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) = \exp\{[y_{ij}\theta_{ij} - \psi(\theta_{ij})]/\phi + c(y_{ij}, \phi)\}$$

and

$$g(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i.$$

Three components specify the GLMM:

- **Distributional assumption:** Choice of the density $f(.|\mathbf{b}_i)$

- **Systematic component:** Linear predictor $\eta_{ij} = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i$

- **Link function** g: $g(\mu_{ij}) = g(E(Y_{ij}|\mathbf{b}_i)) = \mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i$

Generalized linear mixed models (GLMMs)

- Linear mixed models are a special case with f corresponding to the normal density and g to the identity link.

- GLMMs are also useful for non-normal longitudinal responses.

- Estimation is generally harder for GLMMs than for LMMs, see 9.3.

- Often, only a random intercept is considered (in particular for binary responses).

GLMM special cases: logistic regression

- **Distributional assumption:** $Y_{ij}|\mathbf{b}_i \sim \text{Bernoulli}(\mu_{ij})$:

$$
\begin{aligned}
\text{Var}(Y_{ij}|\mathbf{b}_i) &= \text{E}(Y_{ij}|\mathbf{b}_i)(1 - \text{E}(Y_{ij}|\mathbf{b}_i)), \quad \phi = 1 \\
\text{i.e. } v(\mu_{ij}) &= \mu_{ij}(1 - \mu_{ij})
\end{aligned}
$$

- **Systematic component:**

$$
\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i
$$

- **Natural link function:**

$$
\log \frac{P(Y_{ij} = 1|\mathbf{b}_i)}{P(Y_{ij} = 0|\mathbf{b}_i)} = \log \frac{\mu_{ij}}{1 - \mu_{ij}} = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i
$$

GLMM special cases: Poisson regression

- **Distributional assumption:** $Y_{ij}|\mathbf{b}_i \sim \mathsf{Poi}(\mu_{ij})$:

$$
\begin{aligned}
\mathsf{Var}(Y_{ij}|\mathbf{b}_i) &= \mathsf{E}(Y_{ij}|\mathbf{b}_i), \quad \phi = 1 \\
\text{i.e. } v(\mu_{ij}) &= \mu_{ij}
\end{aligned}
$$

- **Systematic component:**

$$
\eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i
$$

- **Natural link function:**

$$
\log(\mathsf{E}(Y_{ij}|\mathbf{b}_i)) = \log(\mu_{ij}) = \eta_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i
$$

The marginal correlation structure

Note that there is no residual covariance $\boldsymbol{R}$ as in the LMM, the Y_{ij} are conditionally independent.

However, the marginal covariance may depend on the covariates:

$$
\begin{aligned}
\mathsf{Cov}(Y_{ij}, Y_{ik}) &= \mathsf{Cov}(\mathsf{E}(Y_{ij}|\boldsymbol{b}_i), \mathsf{E}(Y_{ik}|\boldsymbol{b}_i)) + \mathsf{E}(\mathsf{Cov}(Y_{ij}, Y_{ik}|\boldsymbol{b}_i)) \\
&= \mathsf{Cov}(g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i), g^{-1}(\boldsymbol{x}_{ik}^T\boldsymbol{\beta} + \boldsymbol{z}_{ik}^T\boldsymbol{b}_i)) \\
&\quad + I(j = k)\mathsf{E}(\phi v(g^{-1}(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i))).
\end{aligned}
$$

Overview Chapter 9 - The generalized linear mixed model

Example toenail infections

- Binary reponse severe infection Y_{ij} (yes=1/no=0)

- $n_i = 7$ measurements per subject at $t_{ij} \approx$ 0, 1, 2, 3, 6, 9, 12 months

- Treatment: $T_i = 0$ or $T_i = 1$

Model:
$$Y_{ij}|b_i \sim \text{Bernoulli}(\pi_{ij})$$
with natural link function

$$\text{logit}(\pi_{ij}) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}.$$

and $b_i \sim \mathcal{N}(0, d^2)$.

Interpretation of the parameters for the toenail data

$$\text{logit}(\pi_{ij}) = \log\left(\frac{P(Y_{ij} = 1 | b_i, T_i, t_{ij})}{P(Y_{ij} = 0 | b_i, T_i, t_{ij})}\right) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}.$$

$$\log\left\{\left(\frac{P(Y_{ij} = 1 | b_i, T_i = 1, t_{ij})}{P(Y_{ij} = 0 | b_i, T_i = 1, t_{ij})}\right) \Big/ \left(\frac{P(Y_{ij} = 1 | b_i, T_i = 0, t_{ij})}{P(Y_{ij} = 0 | b_i, T_i = 0, t_{ij})}\right)\right\} = \beta_1 + \beta_3 t_{ij}$$

- β_1 is the log-odds ratio for differences in treatment groups at baseline for a given subject.

- β_3 is the log-odds ratio for differences in the changes per month between treatment groups for a given subject.

The parameters differ from marginal (population average) log-odds ratios!

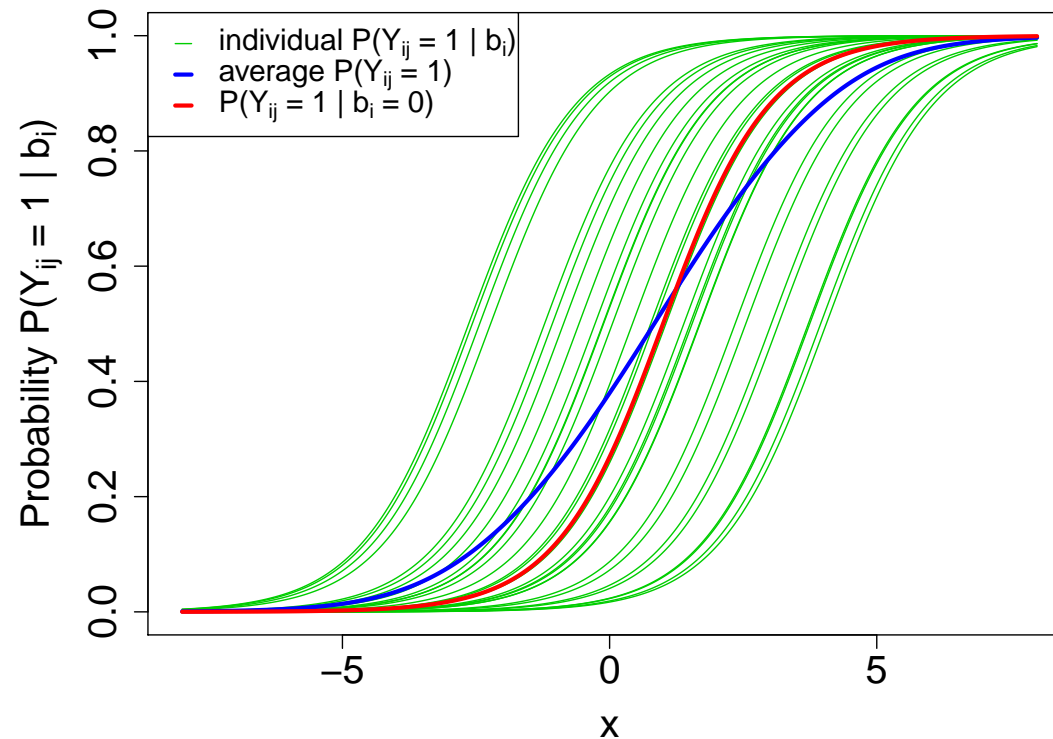# Interpretation of the parameters: Special case logistic model

$$P(Y_{ij} = 1|\mathbf{b}_i) = E(Y_{ij}|\mathbf{b}_i) = \frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right) + 1}$$

Marginal expectation:

$$
\begin{aligned}
P(Y_{ij} = 1) = E(Y_{ij}) \;&=\; E(E(Y_{ij}|\mathbf{b}_i)) \\[2mm]
&=\; E\left[\frac{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right)}{\exp\left(\mathbf{x}_{ij}^T\boldsymbol{\beta} + \mathbf{z}_{ij}^T\mathbf{b}_i\right) + 1}\right] \\[2mm]
&\neq\; \frac{\exp(\mathbf{x}_{ij}^T\boldsymbol{\beta})}{1 + \exp(\mathbf{x}_{ij}^T\boldsymbol{\beta})}
\end{aligned}
$$

$\rightarrow$ Difference to the LMM, where $E(Y_{ij}) = E(E(Y_{ij}|\mathbf{b}_i)) = \mathbf{x}_{ij}^T\boldsymbol{\beta}$

Example for a single continuous x

"Nonlinear function of $E(\boldsymbol{b}_i) \neq E$(nonlinear function of $\boldsymbol{b}_i$)."

Interpretation of the parameters in the GLMM

With link-function $h^{-1}(x) = x$:

$$
\begin{aligned}
E(Y_{ij}) &= \int E(Y_{ij}|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \int (\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \mathbf{x}_{ij}^T \boldsymbol{\beta} \int f(\mathbf{b}_i) d\mathbf{b}_i + \mathbf{z}_{ij}^T \int \mathbf{b}_i f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \mathbf{x}_{ij}^T \boldsymbol{\beta} \\
&= h(\mathbf{x}_{ij}^T \boldsymbol{\beta})
\end{aligned}
$$

With non-linear link-function $h^{-1}(x) \neq x$:

$$
\begin{aligned}
E(Y_{ij}) &= \int E(Y_{ij}|\mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
&= \int h(\mathbf{x}_{ij}^T \boldsymbol{\beta} + \mathbf{z}_{ij}^T \mathbf{b}_i) f(\mathbf{b}_i) d\mathbf{b}_i \\
&\neq h(\mathbf{x}_{ij}^T \boldsymbol{\beta})
\end{aligned}
$$

Interpretation of the parameters in the GLMM

GLMM fixed effects thus differ from the corresponding marginal parameters.

Which of the two quantifies the covariate effects (better)? They both do, although they address slightly different questions.

For example, consider a treatment effect in a logistic model for a disease.

- The parameter in the mixed model quantifies the expected change in the log-odds of the disease for a given individual treated with this treatment. This is of interest, for example, to a doctor treating a given patient.

- The marginal parameter quantifies the expected change in the log-odds of the disease in the population if everyone were given this treatment. This is of interest, for example, to epidemiologists studying the benefit of the treatment in the general population.

Overview Chapter 9 - The generalized linear mixed model

9.1 Introduction to GLMMs

9.2 Interpretation of the parameters

9.3 Estimation and inference

9.4 Implementation and examples

ML estimation

Likelihood contribution for subject i:

$$f(\mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \phi) = \int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\boldsymbol{D}) d\mathbf{b}_i$$

Likelihood:

$$
\begin{aligned}
L(\boldsymbol{\beta}, \boldsymbol{D}, \phi) &= \prod_{i=1}^{N} f(\mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \phi) \\
&= \prod_{i=1}^{N} \int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i
\end{aligned}
$$

ML estimation

For the linear mixed model, the marginal density $f(\mathbf{y}_i|\boldsymbol{\beta}, \boldsymbol{D}, \phi)$ could be obtained explicitly as the $\mathcal{N}(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \boldsymbol{\Sigma}_i)$ density.

For general GLMMs, maximization is made difficult by the product of the N q-dimensional integrals over the random effects. In general, numerical approximations have to be used. This is still an area of active research.

There are three different main approaches:

- approximate the data

- approximate the integrand

- approximate the integral.

Approximate the data: Penalized quasi-likelihood (PQL)

PQL decomposes Y_{ij} into the mean $\mu_{ij} = \mathsf{E}(Y_{ij}|\boldsymbol{b}_i)$ plus an error term ϵ_{ij} with variance $\mathsf{Var}(Y_{ij}|\boldsymbol{b}_i)$. It then uses a Taylor approximation around current estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{b}}_i$.

$$
\begin{aligned}
Y_{ij} &= \mu_{ij} + \varepsilon_{ij} = h(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{b}_i) + \epsilon_{ij} \\
&\approx h(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + \boldsymbol{z}_{ij}^T\widehat{\boldsymbol{b}}_i) + h'(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + \boldsymbol{z}_{ij}^T\widehat{\boldsymbol{b}}_i)\boldsymbol{x}_{ij}^T(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \\
&\qquad\qquad\qquad + h'(\boldsymbol{x}_{ij}^T\widehat{\boldsymbol{\beta}} + \boldsymbol{z}_{ij}^T\widehat{\boldsymbol{b}}_i)\boldsymbol{z}_{ij}^T(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i) + \epsilon_{ij} \\
&= \widehat{\mu}_{ij} + v(\widehat{\mu}_{ij})\boldsymbol{x}_{ij}^T(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + v(\widehat{\mu}_{ij})\boldsymbol{z}_{ij}^T(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i) + \epsilon_{ij}
\end{aligned}
$$

assuming a natural link function, i.e. $h(\cdot) = \psi'(\cdot) \Rightarrow h'(\eta_{ij}) = v(\mu_{ij})$.

In matrix-vector notation with $\widehat{\boldsymbol{\mu}}_i = (\widehat{\mu}_{i1}, \ldots, \widehat{\mu}_{in_i})$, $\widehat{\boldsymbol{V}}_i = \text{diag}(v(\widehat{\mu}_{i1}), \ldots, v(\widehat{\mu}_{in_i}))$:

$$
\boldsymbol{Y}_i \approx \widehat{\boldsymbol{\mu}}_i + \widehat{\boldsymbol{V}}_i \boldsymbol{X}_i(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \widehat{\boldsymbol{V}}_i \boldsymbol{Z}_i(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i) + \boldsymbol{\epsilon}_i
$$

$$
\Leftrightarrow \quad \widehat{\boldsymbol{V}}_i^{-1}(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i) + \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i \approx \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{\epsilon}_i
$$

$$
\Leftrightarrow \quad \boldsymbol{Y}_i^* \approx \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i^*, \tag{9.1}
$$

where $\boldsymbol{Y}_i^*$ denotes the left-hand side of the equation and $\boldsymbol{\varepsilon}_i^* = \widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{\varepsilon}_i$ has mean zero and covariance $\phi\widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{V}_i\widehat{\boldsymbol{V}}_i^{-1} \approx \phi\widehat{\boldsymbol{V}}_i^{-1}$.

This can be viewed as a linear mixed model for the pseudo data $\boldsymbol{Y}_i^*$.

PQL Algorithm

Initialize e.g. $\widehat{\boldsymbol{\beta}}^{[1]}$ as an estimate from a GLM and $\widehat{\boldsymbol{b}}^{[1]} = \boldsymbol{0}$.
For $k = 1, 2, \ldots$ iterate the following two steps until convergence:

- **Step A**: For given $\widehat{\boldsymbol{\beta}}^{[k]}, \widehat{\boldsymbol{b}}^{[k]}$, obtain new pseudo data $\boldsymbol{Y}_i^{*[k]}$ and estimates $\widehat{\boldsymbol{V}}_i^{[k]} = \text{diag}(v(\widehat{\mu}_{i1}^{[k]}), \ldots, v(\widehat{\mu}_{in_i}^{[k]}))$.

- **Step B**: For given pseudo data $\boldsymbol{Y}_i^{*[k]}$, fit model (9.1) and update estimates $\widehat{\boldsymbol{\beta}}^{[k+1]}, \widehat{\boldsymbol{D}}^{[k+1]}$ and $\widehat{\phi}^{[k+1]}$, and predictions $\widehat{\boldsymbol{b}}^{[k+1]}$.

(Breslow & Clayton, 1993)

PQL: Another justification

In the LMM, Henderson's mixed model equations for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ result when maximizing the penalized log-likelihood

$$\log f(\boldsymbol{y}|\boldsymbol{b}) - \frac{1}{2}\boldsymbol{b}^T\boldsymbol{G}^{-1}\boldsymbol{b},$$

where $\boldsymbol{G} = \text{diag}(\boldsymbol{D}, \ldots, \boldsymbol{D})$. Maximizing this penalized log-likelihood for given $\boldsymbol{D}$ and ϕ in the GLMM using Fisher scoring or Newton-Raphson also leads to (9.1).

PQL: Remarks

- PQL is the default in the `gamm()` function in the R package `mgcv`.

- PQL uses an approximate likelihood and is exact only for LMMs. It works better for larger n_i and for Y_{ij} closer to normal, e.g. for larger means in Poisson regression, or larger denominators in binomial proportions. It can be seriously biased for binary responses with relatively small n_i.

- The PQL estimates are consistent only if both $N, n_i \to \infty$.

- REML can be used instead of ML when estimating $\boldsymbol{D}$ and ϕ in (9.1).

- PQL can also be used to give starting values for other methods.

Approximate the integrand: The Laplace approximation

Idea: In

$$\mathcal{I}_i := \int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b}_i, \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{D}) d\boldsymbol{b}_i =: \int_{\mathbb{R}^q} \exp\{Q_i(\boldsymbol{b}_i)\} d\boldsymbol{b}_i$$

approximate $Q_i(\boldsymbol{b}_i)$ by a second-order Taylor approximation around the maximizer $\widehat{\boldsymbol{b}}_i$ of $Q_i(\cdot)$,

$$Q_i(\boldsymbol{b}_i) \approx Q_i(\widehat{\boldsymbol{b}}_i) + \frac{1}{2}(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i)^T \; Q_i''(\widehat{\boldsymbol{b}}_i)(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i).$$

As the normal density integrates to 1,

$$
\begin{aligned}
\mathcal{I}_i &\approx \exp\{Q_i(\widehat{\boldsymbol{b}}_i)\} \int\limits_{\mathbb{R}^q} \exp\left\{\frac{1}{2}(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i)^T Q_i''(\widehat{\boldsymbol{b}}_i)(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i)\right\} d\boldsymbol{b}_i \\
&= (2\pi)^{q/2} |-Q_i''(\widehat{\boldsymbol{b}}_i)|^{-\frac{1}{2}} \exp\{Q_i(\widehat{\boldsymbol{b}}_i)\}, \tag{9.2}
\end{aligned}
$$

where we can derive $-Q_i''(\boldsymbol{b}_i) = \boldsymbol{D}^{-1} + \phi^{-1}\boldsymbol{Z}_i^T \boldsymbol{V}_i \boldsymbol{Z}_i$ with $\boldsymbol{V}_i = \mathrm{diag}(v(\mu_{i1}), \ldots, v(\mu_{in_i}))$.

The modes $\widehat{\boldsymbol{b}}_i$ depend on the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{D}$ and ϕ. The numerical maximization of the likelihood thus iterates between

a) updating the $\widehat{\boldsymbol{b}}_i$ as maximizers of $Q_i(\cdot)$ with current estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{D}}$ and $\widehat{\phi}$ (via a penalized iteratively reweighted least squares (P-IRLS) algorithm),

b) updating estimates for $\boldsymbol{\beta}$, $\boldsymbol{D}$ and ϕ maximizing approximation (9.2) to $\prod_{i=1}^{N} \mathcal{I}_i$ with current $\widehat{\boldsymbol{b}}_i$ (numerical optimization).

Laplace approximation: Remarks

- The Laplace approximation tends to be the fastest approximation and is the default in function `glmer` in `lme4`.

- The approximation is only exact in LMMs, where $Q(\boldsymbol{b}_i)$ is quadratic.

- The approximation tends to be better for larger cluster sizes n_i, and for less "discreteness" of the response (e.g. for Poisson rather than logistic regression) (Joe, 2008).

Approximate the integral: Gaussian quadrature

Reminder: We want to maximize

$$\prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i.$$

Idea: If $\phi(\boldsymbol{z})$ is the density of the multivariate standard normal distribution and $f(\boldsymbol{z})$ a known function, we can approximate

$$\int f(\boldsymbol{z}) \phi(\boldsymbol{z}) d\boldsymbol{z} \approx \sum_{l=1}^{Q} w_l f(\boldsymbol{z}_l),$$

where Q is the degree of the approximation and the w_l are appropriately chosen weights.

Approximate the integral: Gaussian quadrature

- The random effects $\mathbf{b}_i$ need to be standardized: $\boldsymbol{\delta}_i = \mathbf{D}^{-1/2}\mathbf{b}_i$

- Then, with $\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{D}^{1/2}\boldsymbol{\delta}_i$, the integral

$$
\begin{aligned}
f(\mathbf{y}_i|\boldsymbol{\beta}, \mathbf{D}, \phi) &= \int \prod_{j=1}^{n_i} f(y_{ij}|\mathbf{b}_i, \boldsymbol{\beta}, \phi) f(\mathbf{b}_i|\mathbf{D}) d\mathbf{b}_i \\
&= \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\delta}_i, \boldsymbol{\beta}, \phi, \mathbf{D}) f(\boldsymbol{\delta}_i) d\boldsymbol{\delta}_i
\end{aligned}
$$

has the form $\int f(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z}$ and can be approximated as $\int f(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z} \approx \sum_{l=1}^{Q} w_l f(\boldsymbol{z}_l)$.

Gaussian quadrature: Remarks

- Precision increases with the number of quadrature points Q (Trade-off computation time $\leftrightarrow$ precision) $\Rightarrow$ Increase Q until there is no change in the estimates.

- There are standard tables / algorithms for z_l and w_l. However, the z_l are not optimized to cluster in regions where $f(z)$ differs strongly from 0.

- Adaptive Gaussian quadrature (e.g. Molenberghs & Verbeke, 2005, p. 274) chooses the z_l more suitably and typically needs (much) fewer z_l. However, it is more time-consuming, as z_l and w_l - which both depend on $\boldsymbol{\beta}$, $\mathbf{D}$ und ϕ - have to be computed in each iteration.

- Adaptive Gaussian quadrature can be chosen as an option in function `glmer` in `lme4` for $q = 1$. It generally is intractable if $q > 2$.

- For $Q = 1$, adaptive Gaussian quadrature is the Laplace approximation.

Inference

- As ML estimation is used, $\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\beta}, \mathcal{I}^{-1})$ for $N \to \infty$ with $\mathcal{I}$ the Fisher information matrix. Wald-, likelihood ratio or score tests can be used with corresponding χ^2 null distributions.

- The goodness of the approximation depends on N and on the goodness of the approximation to the likelihood.

- PQL estimation is based on the likelihood of pseudo data and thus, no likelihood ratio test is possible. Inference is often based on the LMM for the pseudo data. However, $\widehat{\boldsymbol{\beta}}$ is not consistent in general.

- When testing parameters in $\boldsymbol{D}$, the same asymptotic χ^2 mixture distributions apply as in the LMM (boundary of the parameter space).

Overview Chapter 9 - The generalized linear mixed model

9.1 Introduction to GLMMs

9.2 Interpretation of the parameters

9.3 Estimation and inference

9.4 Implementation and examples

Implementation of GLMMs

- Function `glmer` in R package `lme4`: Laplace approximation and adaptive Gaussian quadrature (for $\mathtt{nAGQ} = Q$ and $Q > 1$)

- Function `glmmPQL` in R package `MASS`: PQL (calls `lme` repeatedly). Used internally by `gamm` in R package `mgcv`.

In SAS: PROC GLIMMIX (PQL) or PROC NLMIXED ((adaptive) Gaussian quadrature).

Example: Toenail infections

$$\text{logit}(\pi_{ij}) = \beta_0 + b_i + \beta_1 T_i + \beta_2 t_{ij} + \beta_3 T_i t_{ij}.$$

```
> for (Q in c(1,3,5,10,20,25)){
+   glmm1 <- glmer(Response ~ Month * Treatment + (1 | ID),
+                  family = binomial, nAGQ = Q)
+   print(fixef(glmm1))
+ }
```

	$Q = 1$	$Q = 3$	$Q = 5$	$Q = 10$	$Q = 20$	$Q = 25$
$\widehat{\beta_0}$	-2.51	-2.03	-1.46	-1.64	-1.62	-1.61
$\widehat{\beta_1}$	-0.30	-0.21	-0.13	-0.16	-0.16	-0.16
$\widehat{\beta_2}$	-0.40	-0.40	-0.38	-0.39	-0.39	-0.39
$\widehat{\beta_3}$	-0.14	-0.14	-0.13	-0.14	-0.14	-0.14

Example: Toenail infections

For $Q = 25$:

```
                Estimate Std. Error z value Pr(>|z|)
(Intercept)     -1.61833    0.43329  -3.735 0.000188 ***
Month           -0.39084    0.04433  -8.817  < 2e-16 ***
Treatment       -0.16081    0.58379  -0.275 0.782968
Month:Treatment -0.13672    0.06799  -2.011 0.044337 *
```

Compare this to a fit using PQL (package `mgcv`):

```
> gamm(Response ~ Month * Treatment, random = list(ID = ~ 1),
            family = binomial)
                    Value Std.Error   DF    t-value p-value
X(Intercept)    -0.7452137 0.2455531 1612 -3.034837  0.0024
XMonth          -0.2960060 0.0314941 1612 -9.398770  0.0000
XTreatment      -0.0355670 0.3481161  292 -0.102170  0.9187
XMonth:Treatment -0.1005177 0.0499676 1612 -2.011659  0.0444
```

Example: Epileptic seizures

Model:

$$
\begin{aligned}
\log E(Y_{ij}|\mathbf{b}_i) &= \log T_{ij} + \beta_1 + \beta_2 F_{ij} + \beta_3 B_i + \beta_4 B_i F_{ij} \\
&\quad + b_{i1} + b_{i2} F_{ij} \\
&= \log T_{ij} + (\beta_1 + b_{i1}) + (\beta_2 + b_{i2}) F_{ij} \\
&\quad + \beta_3 B_i + \beta_4 B_i F_{ij}
\end{aligned}
$$

- T_{ij} = length of observation period: $T_{i1} = 8$, $T_{i2} = T_{i3} = T_{i4} = T_{i5} = 2$.

- $B_i = 0$ for placebo, $B_i = 1$ for progabide.

- $F_{ij} = 0$ for baseline, $t_{ij} = 1$ else.

Interpretation?

Example: Epileptic seizures

Assumptions:

- Conditional on $\mathbf{b}_i$, Y_{ij} is Poisson distributed:

$$\text{Var}(Y_{ij}|\mathbf{b}_i) = \text{E}(Y_{ij}|\mathbf{b}_i), \qquad \phi = 1$$

- $\mathbf{b}_i \sim \mathcal{N}_2(\mathbf{0}, \mathbf{D})$

Group	Time	$\log\left(\frac{E(Y_{ij}|\mathbf{b}_i)}{T_{ij}}\right)$
Placebo	Baseline	$\beta_1 + b_{i1}$
	Follow-up	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2})$
Progabide	Baseline	$(\beta_1 + b_{i1}) + \beta_3$
	Follow-up	$(\beta_1 + b_{i1}) + (\beta_2 + b_{i2}) + \beta_3 + \beta_4$

Epileptic seizures: R-code and results

```
> seiz$followup <- (seiz$visit > 0)
> glmm1 <- glmer(count ~ offset(log(weeks)) + followup * group +
                 (1 + followup | id), family = poisson)
```

```
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.071299   0.140267   7.638 2.21e-14 ***
followupTRUE       -0.002394   0.109092  -0.022   0.9825
group               0.049481   0.192716   0.257   0.7974
followupTRUE:group -0.307159   0.150452  -2.042   0.0412 *
```

```
Random effects:
 Groups Name         Variance Std.Dev. Corr
 id     (Intercept)  0.4999   0.7070
        followupTRUE 0.2319   0.4815   0.17
```

Interpretation? Which method is used for fitting?

Epileptic seizures: R-code and results

Higher-order Gaussian quadrature is only available for models with $q = 1$:

```
> glmm2GQ <- glmer(count ~ offset(log(weeks)) + followup * group +
          (1 | id), family = poisson, nAGQ = 20)
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.03259    0.15269   6.763 1.35e-11 ***
followupTRUE         0.11080    0.04689   2.363   0.0181 *
group               -0.02387    0.21067  -0.113   0.9098
followupTRUE:group  -0.10368    0.06505  -1.594   0.1110

> glmm2LA <- glmer(count ~ offset(log(weeks)) + followup * group +
          (1 | id), family = poisson)
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          1.03265    0.15254   6.769 1.29e-11 ***
followupTRUE         0.11080    0.04672   2.371   0.0177 *
group               -0.02385    0.21047  -0.113   0.9098
followupTRUE:group  -0.10368    0.06482  -1.599   0.1097
```

Final remarks

- There are often convergence problems if several random effects are included in the model, in particular for binary responses if n_i is not large.

- For count data, it is often less problematic to include more than a random intercept in the model.

- Several extensions exist, e.g. to GLMMs with serial correlation, non-normal random effects etc.