Overview

Recap: The Exponential Family

A random variable Y follows a distribution from the exponential family, if the density (or probability mass function) is of the form

$$f(y|\eta,\phi) = \exp\{\phi^{-1}[y\eta - \psi(\eta)] + c(y,\phi)\}$$

with known functions $\psi(\cdot)$ and $c(\cdot,\cdot)$.

- η: natural parameter
- ϕ: scale parameter

The moments can be shown to be

- $\mu = \mathrm{E}(Y) = \psi'(\eta)$
- $\sigma^2 = \mathrm{Var}(Y) = \phi\psi''(\eta) = \phi\psi''(\psi'^{-1}(\mu)) =: \phi v(\mu)$

Thus, the variance is related to the mean through the variance function $v(\cdot) = \psi''(\psi'^{-1}(\cdot))$.

Recap: The Exponential Family - Examples

Densities / probability mass functions:

$$\text{Gaussian}: \quad f(y|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$$

$$\text{Binomial}: \quad f(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$$

$$\text{Poisson}: \quad f(y|\lambda) = \frac{\lambda^y}{y!} \exp(-\lambda)$$

Family	η	ϕ	$\psi(\eta)$	$c(y,\phi)$
Gaussian	μ	σ^2	$\frac{1}{2}\eta^2$	$\frac{y^2}{2\phi} - \frac{1}{2}\log(2\pi\phi)$
Binomial	$\log(\frac{\pi}{1-\pi})$	1	$n\log(1+\exp(\eta))$	$\log\binom{n}{y}$
Poisson	$\log(\lambda)$	1	$\exp(\eta)$	$-y!$

Homework: Verify table, and verify $\mathrm{E}(Y) = \psi'(\eta)$, $\mathrm{Var}(Y) = \phi\psi''(\eta)$.

Recap: Generalized Linear Models (GLMs)

For independent outcomes $Y_1, \ldots, Y_n$ and corresponding vectors of p explanatory variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, the generalized linear model assumes

- Each Y_i has a density $f(y_i|\eta_i, \phi)$ from the exponential family.
- $\mathrm{E}(Y_i) = \mu_i = h(\boldsymbol{x}_i'\boldsymbol{\beta})$ for a known function $h(\cdot)$.

$g(\cdot) = h^{-1}(\cdot)$ is called the link function.
Most often, the natural link function is used, $h(\cdot) = \psi'(\cdot)$, and thus the natural parameter is assumed to satisfy a linear model, $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$.

Recap: Generalized Linear Models - Examples

Regression	η	ϕ	μ	$v(\mu)$	$g(\cdot)$
Linear	μ	σ^2	μ	1	$id(\cdot)$
Logistic	$\log(\frac{\pi}{1-\pi})$	1	π	$\mu(1-\mu)$	$\text{logit}(\cdot)$
Poisson	$\log(\lambda)$	1	λ	μ	$\log(\cdot)$

All three examples use the natural link function, although for binary data, the logit link is sometimes replaced by the probit link $\Phi^{-1}(\cdot)$.

Recap: Generalized Linear Models - Inference

Estimation of $\boldsymbol{\beta}$ is usually done using maximum likelihood estimation. For independent observations, the log-likelihood and score equations are

$$\begin{aligned}
\ell(\boldsymbol{\beta}, \phi) &= \frac{1}{\phi}\sum_{i=1}^{n}[y_i\eta_i - \psi(\eta_i)] + \sum_{i=1}^{n} c(y_i, \phi) \\
S(\boldsymbol{\beta}) &= \sum_{i=1}^{n}\frac{\partial \eta_i}{\partial \boldsymbol{\beta}}[y_i - \psi'(\eta_i)] = \sum_{i=1}^{n}\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} v_i^{-1}[y_i - \mu_i] \stackrel{!}{=} \mathbf{0},
\end{aligned}$$

using $\mu_i = \psi'(\eta_i)$ and $v_i = v(\mu_i) = \psi''(\eta_i)$, giving

$$\frac{\partial \mu_i}{\partial \boldsymbol{\beta}} = \psi''(\eta_i)\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = v_i \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}.$$

Solving for $\boldsymbol{\beta}$ usually has to be done numerically, using iteratively reweighted least squares, Newton-Raphson or Fisher scoring. For the ML estimates, the usual asymptotic likelihood theory can be used to obtain e.g. likelihood ratio tests, Wald test, or score tests. When ϕ is not known, it has to be estimated as well.

Generalized Linear Mixed Models (GLMMs)

Now, go back to the longitudinal data setting, where $Y_{ij}, j = 1, \ldots, n_i$ are marginally correlated observations from the same subject i, $i = 1, \ldots, I$. The generalized linear mixed model assumes that

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

$$f_i(y_{ij}|\eta_{ij}, \phi) = \exp\{\phi^{-1}[y_{ij}\eta_{ij} - \psi(\eta_{ij})] + c(y_{ij}, \phi)\}.$$

- The conditional mean $\mu_{ij} = \mathrm{E}(Y_{ij}|\boldsymbol{b}_i)$ can be modeled as

$$g(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i.$$

  With the natural link function, this becomes $\eta_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i$.

- $\boldsymbol{b}_i$ i.i.d. with density $f(\boldsymbol{b}_i)$. Typically, $\boldsymbol{b}_i \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{D}_0)$ is assumed, and we will also do so in the following.

The Marginal Correlation Structure

Note that the implied marginal correlation structure for $\boldsymbol{Y}$ is independent of the distance between time points - there is no residual covariance matrix $\boldsymbol{R}$ to specify, as there was in the linear mixed model.
However, the correlation structure may depend on the covariates. We have

$$\begin{aligned}
\text{Cov}(Y_{ij}, Y_{ik}) &= \text{Cov}(\text{E}(Y_{ij}|\boldsymbol{b}_i), \text{E}(Y_{ik}|\boldsymbol{b}_i)) + \text{E}(\text{Cov}(Y_{ij}, Y_{ik}|\boldsymbol{b}_i)) \\
&= \text{Cov}(\mu_{ij}, \mu_{ik}) + \text{E}(0) \\
&= \text{Cov}(g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i), g^{-1}(\boldsymbol{x}'_{ik}\boldsymbol{\beta} + \boldsymbol{z}'_{ik}\boldsymbol{b}_i)),
\end{aligned}$$

which can be worked out more explicitly for specific cases. Analogously,

$$\begin{aligned}
\text{Var}(Y_{ij}) &= \text{Var}(\text{E}(Y_{ij}|\boldsymbol{b}_i)) + \text{E}(\text{Var}(Y_{ij}|\boldsymbol{b}_i)) \\
&= \text{Var}(g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i)) + \text{E}(\phi v(g^{-1}(\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{b}_i))),
\end{aligned}$$

and in general, the terms containing the covariates don't cancel in

$$\text{Corr}(Y_{ij}, Y_{ik}) = \text{Cov}(Y_{ij}, Y_{ik}) \Big/ \sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}.$$

Maximum Likelihood Estimation

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_0, \phi)$, where $\boldsymbol{\theta}_0$ includes all unknown parameters in $\boldsymbol{D}_0$. Let q be the dimension of $\boldsymbol{b}_i$.
The likelihood for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (from the marginal distribution for $\boldsymbol{y}$) then is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\theta}) = \int f(\boldsymbol{y}|\boldsymbol{b}; \boldsymbol{\beta}, \phi) f(\boldsymbol{b}|\boldsymbol{\theta}_0) d\boldsymbol{b} \\ &= \prod_{i=1}^{I} \int_{\mathbb{R}^{\boldsymbol{q}}} \prod_{j=1}^{n_i} f_i(y_{ij}|\boldsymbol{b}_i; \boldsymbol{\beta}, \phi) f(\boldsymbol{b}_i|\boldsymbol{\theta}_0) d\boldsymbol{b}_i \end{aligned}$$

The difficult problem for inference in GLMMs is the presence of the I q-dimensional integrals over the random effects.
For the linear mixed model, the marginal distribution could be worked out analytically (another normal). In general, numerical approximations have to be used. This is still an area of active research. Different approaches either approximate the integrand, approximate the data or approximate the integral.

The Laplace Approximation

The Laplace approximation is a method that approximates the integrand.
The integral we would like to approximate is (for the natural link)

$$\int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f_i(y_{ij}|\boldsymbol{b}_i;\beta,\phi) f(\boldsymbol{b}_i|\boldsymbol{\theta}_0) d\boldsymbol{b}_i$$

$$\propto \int_{\mathbb{R}^q} \exp\left\{\phi^{-1}\sum_{j=1}^{n_i}[y_{ij}(\boldsymbol{x}'_{ij}\beta+\boldsymbol{z}'_{ij}\boldsymbol{b}_i)-\psi(\boldsymbol{x}'_{ij}\beta+\boldsymbol{z}'_{ij}\boldsymbol{b}_i)]-\frac{1}{2}\boldsymbol{b}'_i\boldsymbol{D}_0^{-1}\boldsymbol{b}_i\right\}d\boldsymbol{b}_i$$

$$=: \int_{\mathbb{R}^q} \exp\{Q(\boldsymbol{b}_i)\}d\boldsymbol{b}_i =: \mathcal{I}$$

The Laplace approximation has been designed to approximate integrals $\mathcal{I}$ of this form, where $Q(\cdot)$ is known, unimodal and bounded. Let $\widehat{\boldsymbol{b}}$ be the value that maximizes $Q(\boldsymbol{b})$, and approximate $Q(\boldsymbol{b})$ by a second-order Taylor approximation

$$Q(\boldsymbol{b}) \approx Q(\widehat{\boldsymbol{b}}) + \frac{1}{2}(\boldsymbol{b}-\widehat{\boldsymbol{b}})' \; Q''(\widehat{\boldsymbol{b}})(\boldsymbol{b}-\widehat{\boldsymbol{b}}).$$

The Laplace Approximation

This give us, using that the normal density integrates to 1,

$$\begin{aligned}\mathcal{I} &\approx \exp\{Q(\widehat{\boldsymbol{b}}_i)\}\int_{\mathbb{R}^{\boldsymbol{q}}}\exp\left\{\frac{1}{2}(\boldsymbol{b}_i-\widehat{\boldsymbol{b}}_i)'Q''(\widehat{\boldsymbol{b}}_i)(\boldsymbol{b}_i-\widehat{\boldsymbol{b}}_i)\right\}\boldsymbol{b}_i\\ &= (2\pi)^{q/2}|-Q''(\widehat{\boldsymbol{b}}_i)|^{-\frac{1}{2}}\exp\{Q(\widehat{\boldsymbol{b}}_i)\},\end{aligned}$$

where

$$\begin{aligned}-Q''(\boldsymbol{b}_i) &= \boldsymbol{D}_0^{-1}+\phi^{-1}\sum_{j=1}^{n_i}\boldsymbol{z}_{ij}\psi''(\boldsymbol{x}'_{ij}\boldsymbol{\beta}+\boldsymbol{z}'_{ij}\boldsymbol{b}_i)\boldsymbol{z}'_{ij}\\ &= \boldsymbol{D}_0^{-1}+\phi^{-1}\boldsymbol{Z}'_i\boldsymbol{V}_i\boldsymbol{Z}_i \quad \text{with} \quad \boldsymbol{V}_i=\text{diag}(\text{Var}(Y_{ij}|\boldsymbol{b}_i)).\end{aligned}$$

The mode $\widehat{\boldsymbol{b}}_i$ depends on the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. The numerical maximization of the likelihood will thus iterate between updating $\widehat{\boldsymbol{b}}_i$ and $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})$.

The Laplace Approximation

- The Laplace approximation tends to be the fastest approximation and is the default implemented in the R package `lme4`, function `glmer`.
- The approximation is only exact in special cases such as the linear mixed model, where $Q(\boldsymbol{b})$ is actually a quadratic function in $\boldsymbol{b}$.
- The approximation tends to be better for larger cluster sizes n_i, and for less "discreteness" of the response (e.g. it is better for Poisson than for logistic regression) (Joe, 2008).
- The approximation can be improved by adding higher-order terms to the Taylor approximation.

Penalized Quasi-Likelihood (PQL)

PQL is a method that approximates the data by the mean $E(Y_{ij}|\boldsymbol{b}_i)$ plus an error term ε_{ij} with variance equal to $\text{Var}(Y_{ij}|\boldsymbol{b}_i)$. It then uses a Taylor approximation.

$$\begin{aligned} Y_{ij} &\approx \mu_{ij} + \varepsilon_{ij} = h(\boldsymbol{x}'_{ij}\beta + \boldsymbol{z}'_{ij}\boldsymbol{b}_i) + \varepsilon_{ij} \\ &\approx h(\boldsymbol{x}'_{ij}\widehat{\beta} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b}}_i) + h'(\boldsymbol{x}'_{ij}\widehat{\beta} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b}}_i)\boldsymbol{x}'_{ij}(\beta - \widehat{\beta}) \\ &\qquad + h'(\boldsymbol{x}'_{ij}\widehat{\beta} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b}}_i)\boldsymbol{z}'_{ij}(\boldsymbol{b}_i - \widehat{\boldsymbol{b}}_i) + \varepsilon_{ij} \end{aligned}$$

This gives (assuming a natural link function, i.e. $h(\cdot) = \psi'(\cdot)$)

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

where $\widehat{\boldsymbol{\mu}}_i$ contains the values $\widehat{\mu}_{ij} = h(\boldsymbol{x}'_{ij}\widehat{\beta} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b}}_i)$, $\widehat{\boldsymbol{V}}_i$ is the diagonal matrix with elements $v(\widehat{\mu}_{ij}) = h'(\boldsymbol{x}'_{ij}\widehat{\beta} + \boldsymbol{z}'_{ij}\widehat{\boldsymbol{b}}_i)$, and $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ contain the $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_{ij}$, respectively.

Penalized Quasi-Likelihood (PQL)

Reordering and pre-multiplication with $\widehat{\boldsymbol{V}}_i^{-1}$ gives

$$\boldsymbol{Y}_i^* := \widehat{\boldsymbol{V}}_i^{-1}(\boldsymbol{Y}_i - \widehat{\boldsymbol{\mu}}_i) + \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i \approx \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i^*, \qquad (1)$$

where $\boldsymbol{\varepsilon}_i^* = \widehat{\boldsymbol{V}}_i^{-1}\boldsymbol{\varepsilon}_i$ still has mean zero.
This can be viewed as a linear mixed model for the pseudo data $\boldsymbol{Y}_i^*$.

$\Rightarrow$ **Algorithm**

- Step A: For given values of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, empirical Bayes estimates for $\boldsymbol{b}_i$ and resulting pseudo data $\boldsymbol{Y}_i^*$ are computed.
- Step B: For given pseudo data $\boldsymbol{Y}_i^*$, model (1) is fit and estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are updated.

Iterate Steps A and B until convergence.

Penalized Quasi-Likelihood (PQL)

Another justification for PQL: In the linear mixed model, when obtaining the BLUPs $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ for given $\boldsymbol{\theta}$, Henderson's mixed model equations arise from maximizing the penalized log-likelihood

$$\log f(\boldsymbol{y}|\boldsymbol{b}) - \frac{1}{2}\boldsymbol{b}'\boldsymbol{D}^{-1}\boldsymbol{b}.$$

If we maximize this penalized log-likelihood for given $\boldsymbol{\theta}$ for the generalized case, the Fisher's method of scoring or Newton-Raphson both lead to (1) as well.

Penalized Quasi-Likelihood (PQL)

- More details can be found in Breslow & Clayton (JASA, 1993).
- This approach is called penalized quasi-likelihood, as it is based on a quasi-likelihood involving only first and second (conditional) moments, plus a penalty term for the random effects.
- PQL is the default in the `gamm()` function in the R package `mgcv` (for the generalized case), which is based on the `glmmPQL` function in the `MASS` library.
- A refined version called pseudolikelihood (Wolfinger & O'Connell, JSCS, 1993) is implemented in the `SAS` macro `GLIMMIX` (incorporates overdispersion).
- PQL uses an approximate likelihood and is exact only for LMMs. It works better the closer the Y_i are to normal. Thus, it works better for larger means in Poisson regression, or larger denominators in binomial proportions.
- The PQL estimates are inconsistent. There exist some bias corrections for PQL (Breslow & Lin, Biometrika, 1995, Lin & Breslow, JASA, 1996). Also, the accuracy may be improved by using a second order Taylor approximation.
- REML can be used instead of ML when estimating $\boldsymbol{\theta}$ in model (1).
- PQL can also be used to give reasonable starting values for other methods.

(Adaptive) Gaussian Quadrature

Gaussian Quadrature approximates the integral by a weighted sum.

$$\int f(x)\phi(x)dx \approx \sum_{l=1}^{L} w_l f(x_l)$$

for L nodes or quadrature points $x_l, l = 1, \ldots, L$, and appropriate weights w_l, where $f(x)$ is a known function, and $\phi(x)$ is the (multivariate) normal density.

- For Gaussian Quadrature, the x_l are solutions to the Lth order Hermite polynomial. This is not optimal for a density $f(x)$ from the exponential family.
- Adaptive Gaussian Quadrature chooses the x_l more suitably and typically needs (much) fewer nodes. However, it is more time-consuming, as it has to first calculate the x_l as well as the w_l - which both depend on β and θ and thus need to be updated in each step of the iteration.

Increasing L improves the approximation. Interestingly, for $L = 1$, Adaptive Gaussian Quadrature reduces to the Laplace approximation.

Adaptive Gaussian Quadrature can be chosen as an option in function `glmer` in R package `lme4`. It generally is intractable, however, if $q > 2$.

Alternatives

An alternative is to treat the random effects as missing data and use the Expectation Maximization (EM)-algorithm (Dempster, Laird & Rubin, JRSS-B, 1977). However, computation of the expectation in the E-step is at least as difficult as computation of the (marginal) log-likelihood, involving the ratio of two intractable integrals. Solutions and techniques that have been proposed include the Laplace approximation, Monte Carlo-EM with Metropolis-Hastings algorithm or Gibbs sampler, and importance sampling.

A Bayesian approach puts priors on all model parameters and simulates from the posterior using MCMC. Posterior means give estimates for the model parameters, and the imputed values for the random effects correspond to their BLUPs.

Prediction of the Random Effects

The random effects can be predicted using either the posterior mode or the posterior mean (empirical Bayes estimates). These will be different in general, as the posterior density is generally no longer normal.

The best predictor of $\boldsymbol{b}$ for given $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ (in the sense of minimum mean squared error of prediction $\mathrm{E}(\boldsymbol{b} - \widetilde{\boldsymbol{b}})^2$) is $\rightarrow$ blackboard

$$\widetilde{\boldsymbol{b}} = \mathrm{E}_{(\boldsymbol{\beta},\boldsymbol{\theta})}(\boldsymbol{b}|\boldsymbol{y}),$$

which suggests $\mathrm{E}_{(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})}(\boldsymbol{b}|\boldsymbol{y})$ as a predictor. Note that

$$\widetilde{\boldsymbol{b}} = \int \boldsymbol{b} f(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\theta}) d\boldsymbol{b} = \frac{\int \boldsymbol{b} f(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{\beta},\phi) f(\boldsymbol{b}|\boldsymbol{\theta}_0) d\boldsymbol{b}}{\int f(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{\beta},\phi) f(\boldsymbol{b}|\boldsymbol{\theta}_0) d\boldsymbol{b}}$$

again necessitates numerical integration. As an alternative, statistical software packages often calculate the posterior mode instead of the posterior mean, which maximizes $f(\boldsymbol{b}|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\theta}) \propto f(\boldsymbol{y}|\boldsymbol{b},\boldsymbol{\beta},\phi) f(\boldsymbol{b}|\boldsymbol{\theta}_0)$.

PQL directly estimates $\boldsymbol{b}$, and for MC-EM or MCMC, the means of the generated samples of $\boldsymbol{b}$ also allow to estimate $\boldsymbol{b}$.

Standard errors

In Bayesian methods, standard errors can be derived as part of the MCMC calculations, and these also account for uncertainty in $\boldsymbol{\theta}$.

For PQL, the iterative fitting of linear mixed models can be used to derive approximate standard errors (not accounting for uncertainty in $\boldsymbol{\theta}$) from those in the linear mixed model. For the linear mixed model (1), the BLUPs are

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{b}} \end{pmatrix} \approx (\boldsymbol{C}'\boldsymbol{V}\boldsymbol{C} + \boldsymbol{G})^{-1}\boldsymbol{C}'\boldsymbol{V}\boldsymbol{Y}^*$$

with, as before, $\boldsymbol{C} = (\boldsymbol{X}|\boldsymbol{Z})$, $\boldsymbol{G} = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{pmatrix}$ and $\boldsymbol{V} = \frac{1}{\phi}\text{Var}(\boldsymbol{y}|\boldsymbol{b})$. Thus, e.g.

$$\text{Cov}\left(\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{b}} \end{pmatrix} \middle| \boldsymbol{b}\right) \approx (\boldsymbol{C}'\boldsymbol{V}\boldsymbol{C} + \boldsymbol{G})^{-1}\boldsymbol{C}'\boldsymbol{V}\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{V}\boldsymbol{C} + \boldsymbol{G})^{-1}.$$

Hypothesis Tests for β

The model is fitted using maximum likelihood. Thus, in principle, the same asymptotics apply, $\widehat{\beta}$ is asymptotically normally distributed with mean β and the inverse Fisher information matrix as covariance matrix. Wald, likelihood ratio, or score tests can be used, comparing the test statistic to an appropriate chi-square distribution.

However, the validity depends on the goodness of the approximations to the likelihood used in estimation. Numerical methods are also needed to calculate the observed Fisher information, which is based again on the log-likelihood. Also, be careful to base likelihood ratio tests on the likelihood of the observed data, and not on the likelihood of the pseudo data used in PQL in (1).

Often, inference is based on the linear mixed model used in PQL. While this gives valid precision estimates for the estimates, $\widehat{\beta}$ is in general not consistent. One also has to consider the validity of normal, t- or F-distributions used in tests. In linear mixed models, normality of $\widehat{\beta}$ follows directly from the normality assumption. In the GLMM, we only have normality asymptotically, when $I \to \infty$. For the empirical Bayes estimates, normality might be questionable.

Hypothesis Tests for D

For the longitudinal generalized linear mixed model, the same asymptotics for testing variance/covariance parameters apply as for the longitudinal linear mixed model (boundary!). Remember that likelihood ratio tests, Wald tests or Score tests can be used, where the latter two involve additional derivatives and a constraint optimization problem to derive the one-sided versions (Molenberghs & Verbeke, AmStat, 2007).

Additional Resources

Remember that the generalized linear mixed model also allows you to include smooth functions modeled using penalized splines. The typical approach is to maximize a penalized log-likelihood

$$\log f(\boldsymbol{y}|\boldsymbol{b}) - \frac{1}{2\lambda}\boldsymbol{b}'\boldsymbol{b},$$

where $\frac{1}{2\lambda}\boldsymbol{b}'\boldsymbol{b}$ is a penalty term imposing a smoothness constraint on the estimated function. This is then equivalent to PQL estimation in a generalized linear mixed model where $\boldsymbol{Y}|\boldsymbol{b}$ follows an exponential family distribution with $g(\mathrm{E}(\boldsymbol{Y}|\boldsymbol{b})) = \boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b}$, and $\boldsymbol{b} \sim N(\boldsymbol{0}, \lambda \boldsymbol{I}_r)$.
You can read more about this approach in Ruppert, Wand & Carroll, 2003. While this is a book on "Semiparametric Regression", it actually also has nice short introductions to LMMs and GLMMs.

Another resource on Generalized Additive Mixed Models (GAMMs) is "Generalized Additive Models" (Wood, 2006), which will give you more background also on the implementation in the R package `mgcv` and its function `gamm()`.

Another resource: "Generalized, Linear, and Mixed Models" (McCulloch & Searle, 2004).

Generalized Linear Mixed Models in R

See `examples_mixed_models.R` for example code for

- simulating data that follows a generalized linear mixed model.
- fitting a generalized linear mixed model.

For examples of Generalized Linear Mixed Models in SAS, see Molenberghs & Verbeke, 2006.