
Generalized Linear Mixed Models

Peter Dalgaard

Department of Biostatistics
University of Copenhagen

Mixed Models in R, January 2006

Overview of GLMM

- ▶ Generalized Linear Models
- ▶ Mixed effects extensions
- ▶ Approximations
- ▶ Examples

Generalized Linear Models

- ▶ Statistical distribution (exponential) family
- ▶ Link function transforming mean to linear scale
- ▶ Deviance
- ▶ Examples; Binomial, Poisson, Gaussian (σ known — in principle)
- ▶ Canonical link functions: logit, log, identity
- ▶ Fit using `glm` in R

Generalized Linear Mixed Models

- ▶ Basic idea: Just add random effects on the linear scale
- ▶ Prototypically: Logistic regression with random effects

$$\log(p/(1 - p)) = X_i\beta + Z_i b_i$$

(Notice that there is no error term because we are modeling probabilities)

- ▶ Typically, a normal distribution is assumed for the random effects.
- ▶ Notice that the *interpretation* can be quite different from a GEE analysis.

Problem: The Likelihood

The likelihood of a GLMM model will involve an integral, which cannot in general be computed explicitly. Following Molenberghs and Verbeke, this contribution has the form

$$\int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|b_i, \beta, \phi) f(b_i|D) db_i$$

They proceed to explain the main approximations as

- ▶ Approximating the integrand (Laplace)
- ▶ Approximating the data (PQL)
- ▶ Approximating the integral

The Laplace Method

- ▶ A trick from Bayes theory:
- ▶ Likelihoods often look like Gaussian densities, which have known integrals
- ▶ The log-likelihood is then close to parabolic in shape
- ▶ Approximate using value at peak and Hessian (comes automatically from optimizers)
- ▶ Apply to the joint likelihood of (y_i, b_i)

The PQL approach

A Taylor expansion of

$$Y_{ij} = h(x'_{ij}\beta + z'_{ij}b_i) + \varepsilon_{ij}$$

yields (in vector notation)

$$\hat{V}_i^{-1}(Y_i - \hat{\mu}_i) + X_i\hat{\beta} + Z_i\hat{b}_i \approx X_i\beta + Z_ib_i + \varepsilon_i^*$$

which gives a straightforward updating scheme, considering the LHS as *pseudo data*.

This is known as *penalized quasi likelihood* because it obtains from optimizing a quasi-likelihood (involving only 1st and 2nd derivatives) with a penalty term on the random effects.

Adaptive Gauss Quadrature

This approximates the integral by evaluating it at a set of *quadrature point* and obtaining an approximation of the form

$$\sum w_i f(x_i)$$

Optimal weights and x_i have been tabulated for expansions in terms of Hermite polynomials.

The “adaptive” refers to scaling the integrand using the Hessian at the optimum point, like in the Laplace method.

Drawback: Complexity increases exponentially with dimension of random effects vector.

R Functions for GLMM

- ▶ `lmer` used to be able to handle all three approaches, but currently cannot do AGQ
- ▶ Model specification is quite trivial: Just add a `family` argument
- ▶ Older PQL code is `glmmPQL` in the `MASS` package (based on `lme`)

Example 1: Smoking in School Children

```
library(foreign)
library(lme4)
dd <- read.xport("~/niss.xpt")
names(dd) <- tolower(names(dd))
dd$ryger <- dd$rygvn<2
(fit <- lmer(ryger~koen + alder + factor(klastrin) + factor(skole)
  + (1|skole)+(1|klassenr), family=binomial, data=dd))
anova(fit)
waldtest <- function(fit, ix) {x <- fixef(fit)[ix];
  V <- as.matrix(vcov(fit))[ix, ix]; x%*%solve(V, x)}
waldtest(fit, 4:5)
```

Example 2: Pressure sensitivity of teeth

(Misc. data manipulation skipped)

```
d$response <- c(.5,1,0)[d$hilo]
d$type <- gl(2,240,labels=c("Implant","Control"))
d$dif <- with(d,test-ref)
```

```
library(lme4)
fit <- lmer(response~dif*type +
            (dif-1 |ID) + (dif-1|ID:type),
            data=d,family=quasibinomial)
```

```
ff <- plogis(fitted(fit))
plot(ff,d$response-ff)
lines(smooth.spline(ff, d$response-ff))
plot(ff,(d$response-ff)^2)
lines(smooth.spline(ff, (d$response-ff)^2))
```