TMA4315 Generalized linear models H2018
Module 8: Generalized linear mixed effects models (GLMM)

Mette Langaas and Ingeborg Hem, Department of
Mathematical Sciences, NTNU

15.11.2018 [PL], 16.11.2018 [IL]

(Latest changes: 16.11: IL added links, 13.11.2018 - first version).

Overview

Aim: Present methods for analysing correlated responses in a generalized linear models setting - LMM meets GLM to become *GLMM*.

Also here (as in module LMM) we will only consider *two-level models* and in particular focus on *random intercept models* for binomial and Poisson responses. Emphasis will be on understanding.

Learning material
- Textbook: Fahrmeir et al (2013): Chapter 7.5: 389-394, 7.7.
- Classnotes (15.11.2018)

Topics
- beaches example - revisited
- notation
- the generalized linear mixed effect model (three ingredients)
- the GLMM with random intercept
- the marginal model
- parameter estimation and Laplace approximation
- summing up: what do we need to know about the GLMM?
- additional info on different software (not on reading list)

Beaches example - revisited

This example is taken from Zuur et al. (2009, chapter 5, pages 101-142), and data are referred to as RIKZ. Data were collected on nine different beaches in the Netherlands with the aim to investigate if there is a relationship between:

▶ richness of species (number of species observed) and
▶ NAP: the height of the sampling station compared to mean tidal level.

Data: 45 observations, taken at 9 beaches:

▶ beach: the beach that the samples were taken, for each beach 5 different samples were taken.
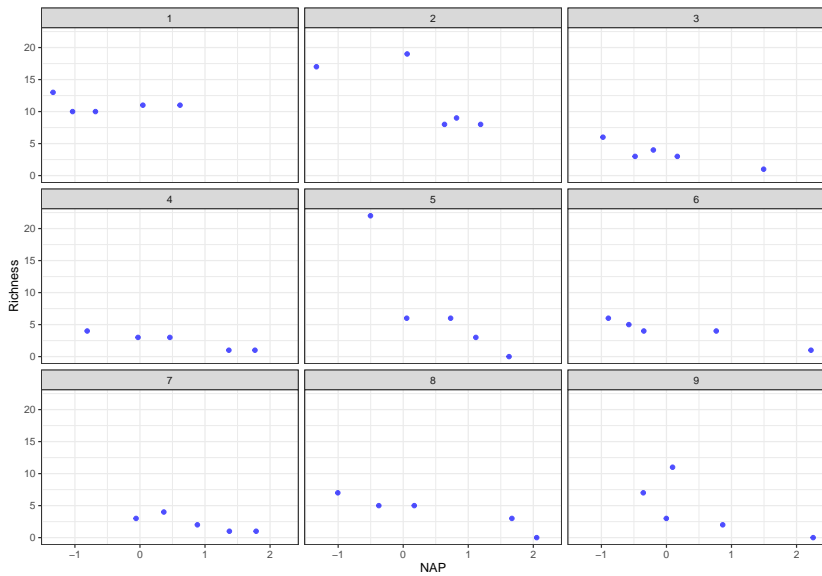
We want to use the richness of species as the response in a regression model.

In Module 7 we assumed that the response could be viewed as a normal variable, but now we assume that the response comes from a *Poisson distribution*.

As in the LMM module, we could (1) ignore the beach effect, (2) include it as a fixed effect, or (3) use a random effect. Here we will use (3), and extend the GLM framework to include random effects

But - first: what do we remember about GLMs (and Poission regression with log-link in particular)?

Assumptions for GLM model (Poisson):

1. Random component from exponential family: We have for our beaches: $Y_i \sim \text{Poisson}(\lambda_i)$, with $\mathsf{E}(Y_i) = \lambda_i$, and $\text{Var}(Y_i) = \lambda_i$.

2. Linear predictor: $\eta_i = \mathbf{x}_i^T \beta$.

3. Log link

$$\eta_i = \ln(\lambda_i) = g(\lambda_i)$$

and (inverse thereof) response function

$$\lambda_i = \exp(\eta_i)$$

Canonical link for Poisson is log.

Generalized linear mixed effects models

Notation

combining GLM (Module 5) and LMM (Module 7)

We have n_i repeated observations ($j = 1, \dots, n_i$) from each of $i = 1, \dots, m$ clusters (or individuals).

▶ Responses: $Y_{i1}, Y_{i2}, \dots, Y_{in_i}$ (e.g. species richness at beach i sample j)

▶ Covariates: $\mathbf{x}_{i1}, \mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}$ (e.g. NAP for beach i sample j)

The covariates $\mathbf{x}_{ij}$ are $p \times 1$ vectors (as before, k covariates and one intercept so $p = k + 1$).

Repetition GLM

1. **Random component** - exponential family (now y_{ij})

$$f(y_{ij} \mid \theta_{ij}) = \exp\left(\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\phi} \cdot w_{ij} + c(y_{ij}, \phi, w_{ij}) \right)$$

▶ θ_{ij} is called the canonical parameter and is a parameter of interest
▶ ϕ is called a nuisance parameter (and is not of interest to us=therefore a nuisance (plage))
▶ w_{ij} is a weight function, in most cases $w_{ij} = 1$
▶ b and c are known functions.

We looked in detail into the binomial, Poisson and gamma distribution (in addition to the normal).

2. **Linear predictor** (previously "systematic component"):

$$\eta_{ij} = \mathbf{x}_{ij}^T \beta$$

3. **Link function** g- **and response function** h

$$\eta_{ij} = g(\mu_{ij})$$

$$\mu_{ij} = h(\eta_{ij})$$

we add a random component (effect) to the *linear predictor*!

2.

$$\eta_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{u}_{ij}^T \gamma_i$$

and we make the same assumptions as for the LMM

$$\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$$

and independent from covariates. Observe: we still assume normality for the random effects!

In addition, the mean $\mu_{ij} = \mathsf{E}(Y_{ij} \mid \gamma_i)$ is now to be seen to be conditional on the random effect γ_i. This means that the specified response distribution is conditional on the random effect: $f(y_{ij} \mid \gamma_i)$.

GLMM model assumptions

Distributional assumptions

(previously "1. Random component from exponential family")

- ▶ Given the random effects γ_i and the covariates $\mathbf{x}_{ij}$ and $\mathbf{u}_{ij}$
- ▶ the responses Y_{ij} are *conditionally independent* and
- ▶ the *conditional distribution* $f(y_{ij} \mid \gamma_i)$ belongs to and exponential family.

Structural assumptions

(previously "2. Systematic component" and "3. Link function")

▶ The *conditional mean* $\mu_{ij} = \mathsf{E}(Y_{ij} \mid \gamma_i)$ is linked to

▶ the *linear predictor* $\eta_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{u}_{ij} \gamma_i$

▶ through the link function

$$\eta_{ij} = g(\mu_{ij})$$

▶ or equivalently through the response function

$$\mu_{ij} = h(\eta_{ij})$$

▶ where $h^{-1} = g$ (inverse functions).

Distributional assumptions for random effects

▶ The random effects γ_i, $i = 1, \ldots, m$ are independent and identically distributed

$$\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$$

▶ where the covariance matrix $\mathbf{Q}$ is a $(q+1) \times (q+1)$ positive definite.

▶ An important special case is $\mathbf{Q} = diag(\tau_0^2, \tau_1^2, \ldots, \tau_q^2)$.

Any questions? Notice that we still have normal distribution for the random effects!

Remark: for LMMs we used that $\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, but said that it was possible to let the error terms be correlated $\varepsilon \sim N(\mathbf{0}, \mathbf{R})$ where $\mathbf{R}$ could have non-zero off-diagonal elements. However, for GLMMs this is much more complicated.

Alternative two-step formulation

Consider a random variable Y_{ij} and a random effect γ_i for cluster i.
Alternative two step formulation:

1. $\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$
2. Given the realization of γ_i, then the Y_{ij}s are independent and each have the distribution $f(y_{ij} \mid \gamma_i)$ with mean $\mu_{ij} = h(\eta_{ij})$ and linear predictor $\eta_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{u}_{ij}^T \gamma_i$.

Remark: this conditional formulation is "easy" to write down, but to do inference we need the marginal distribution of all Y_{ij}s together to construct the likelihood. We will soon see that this is hard.

GLMM random intercept model

Distributional assumptions: we will focus on binomial and Poisson responses.

Structural assumptions: The random effect is added to the linear predictor

$$\eta_{ij} = \mathbf{x}_{ij}^T \beta + \gamma_{0i}$$

and we only consider canonical links, so logit for binomial and log for Poisson.

Distributional assumptions for random effects

The γ_{0i}, $i = 1, \ldots, m$ are independent and identically distributed

$$\gamma_{0i} \sim N(0, \tau_0^2)$$

Poisson random intercept model

Distributional assumptions: The conditional distribution of the response Y_{ij} is Poisson with mean λ_{ij}

$$f(y_{ij} \mid \gamma_i) = \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \exp(-\lambda_{ij}) \text{ for } y = 0, 1, 2, ...$$

This is conditional on λ_{ij}, which means that it is really conditional on the linear predictor - since the mean is a function of the linear predictor, and in the linear predictor we have fixed and random effects. So, conditional on the values for the fixed and random effects.

Then the observations Y_{ij} are *conditionally* independent for all i and j - but, not *marginally* independent.

Structural assumptions:

$$\eta_{ij} = \mathbf{x}_{ij}^T \beta + \gamma_{0i}$$

$$\ln(\lambda_{ij}) = \eta_{ij} \text{ or } \lambda_{ij} = \exp(\eta_{ij})$$

Distributional assumptions for random effects

The γ_{0i}, $i = 1, \ldots, m$ are independent and identically distributed

$$\gamma_{0i} \sim N(0, \tau_0^2)$$

Overdispersion

It is not necessary that $n_i > 1$ to use this model, and this model can be used to take care of overdispersion (when all $n_i = 1$). For the Poisson case, this would be equivalent to using a Poisson log-normal distribution

Beach-example: Poisson GLMM with beach as random intercept

(we will talk about parameter estimation soon)

```
RI=glmer(Richness~NAP +(1|Beach),data=RIKZ,
         family=poisson(link=log))
summary(RI)$coefficients
```

```
##               Estimate Std. Error   z value    Pr(>|z|)
## (Intercept)  1.6623311 0.17372603  9.568693 1.082653e-2:
## NAP         -0.5038922 0.07535222 -6.687158 2.275463e-1:
```

```
summary(RI)$varcor
```

```
##  Groups Name        Std.Dev.
##  Beach  (Intercept) 0.47428
```

Beach-example: Poisson GLMM with overdispersion

Sample is 1:nrow(RIKZ)

```
RIod=glmer(Richness~NAP +(1|Beach) +(1|Sample),
           data=RIKZ,family=poisson(link=log))
summary(RIod)$coefficients
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept)  1.636575 0.17475924  9.364739 7.623879e-21
## NAP         -0.551958 0.09671449 -5.707087 1.149261e-08
```
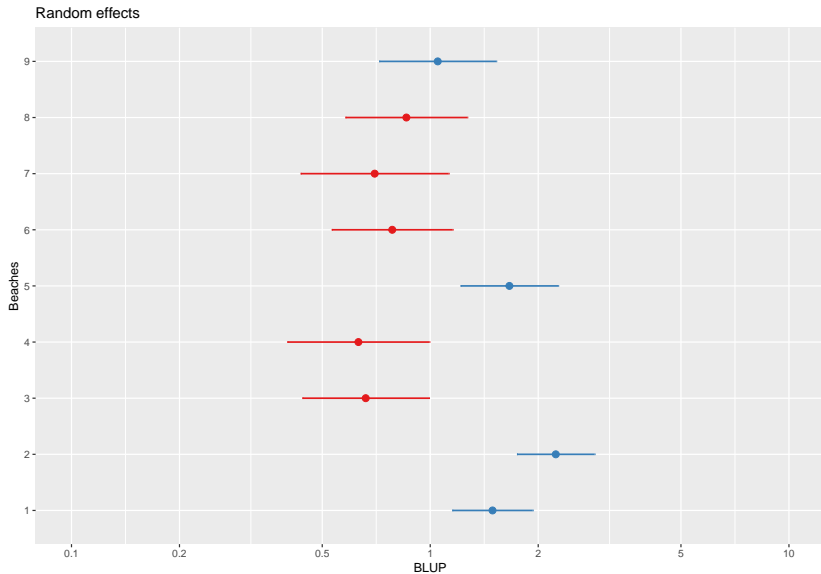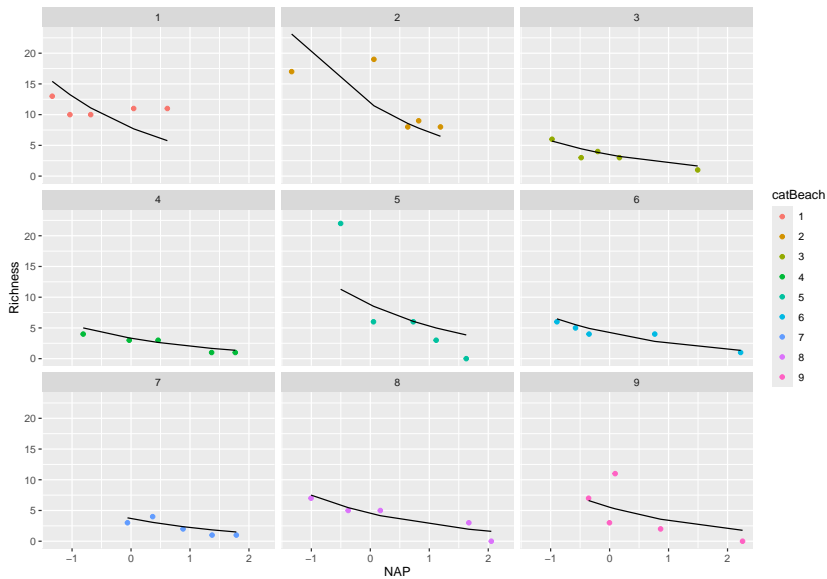
```
summary(RIod)$varcor
```

```
##  Groups Name        Std.Dev.
##  Sample (Intercept) 0.26555
##  Beach  (Intercept) 0.45822
```
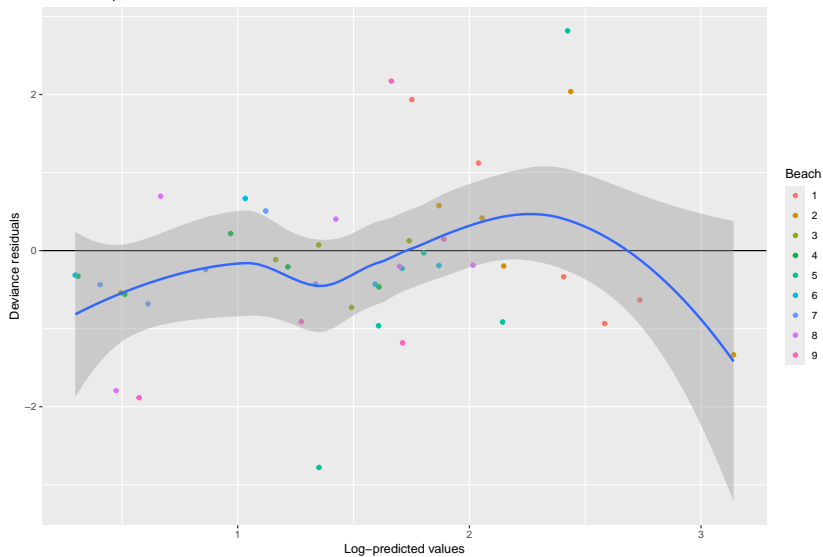
```r
library(sjPlot)
library(ggplot2)

plot_model(RI, type = "re") + ylab("BLUP") + xlab("Beaches")
```
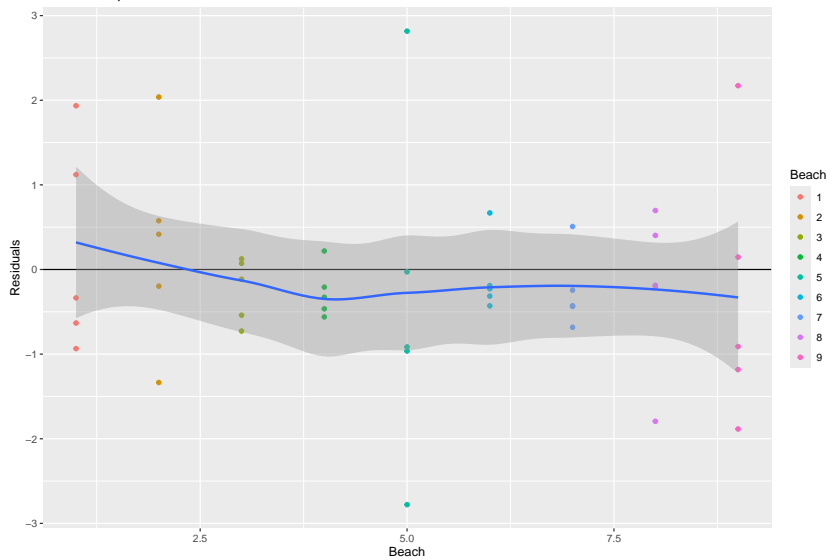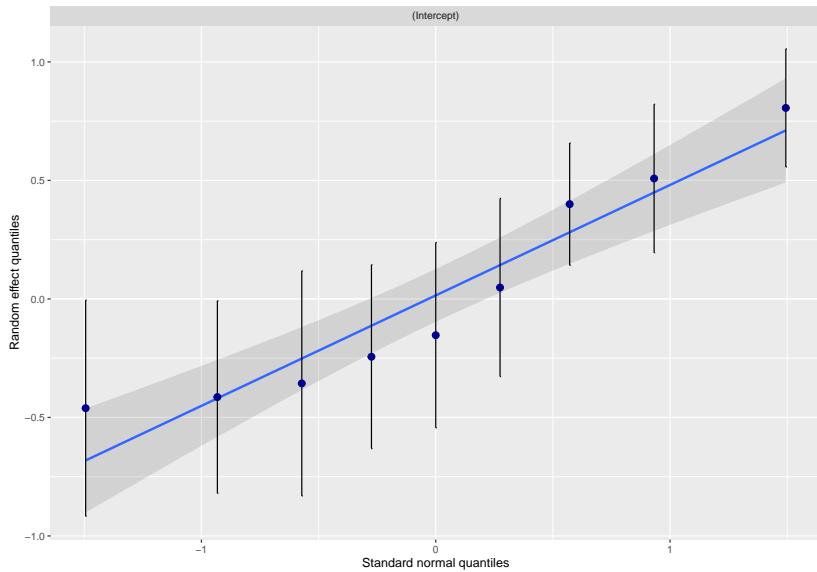
Residuals plot

Relationship between Beach and residuals

Comment on what you see. Would you expect that the solid and dashed curve were identical?

Binomial random intercept model

We will study this model in the interactive session.

Simulation example

Aim: to show the difference between random intercept and random slope logit models, and to compare the mean cluster curve with the population curve (as for the Poisson model).

This example a slightly modified version (added line and changed to ggplot) of an example in the Lecture notes by Magne Aldrin, Norwegian Computing Centre (which is way the notation differs slightly from what we have used).

The model we use is

$$\begin{pmatrix} \gamma_{0i} \\ \gamma_{1i} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

and then

$$\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i}$$

for the random intercept model and

$$\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_{0i} + \gamma_{1i} x_{ij}$$

for the random slope model where $\beta_0 = 3$ and $\beta_1 = 1$. Logit-link is used so $\eta_{ij} = \ln(\frac{\pi_{ij}}{1-\pi_{ij}})$.

Only data for the random intercept and slope are simulated, while parameter estimates are considered known. The number of clusters is $m = 1000$ and $n_i = 201$ value of x from -10 to 10 are studied.

Below 30 random lines are plotted together (in different colours). The black dashed line is $\exp(\beta_0 + \beta_i x)/(1 + \exp(\beta_0 + \beta_i x))$ (denoted "not mean" in the plot) and the black solid line is the population mean (really, the average of the m curves).
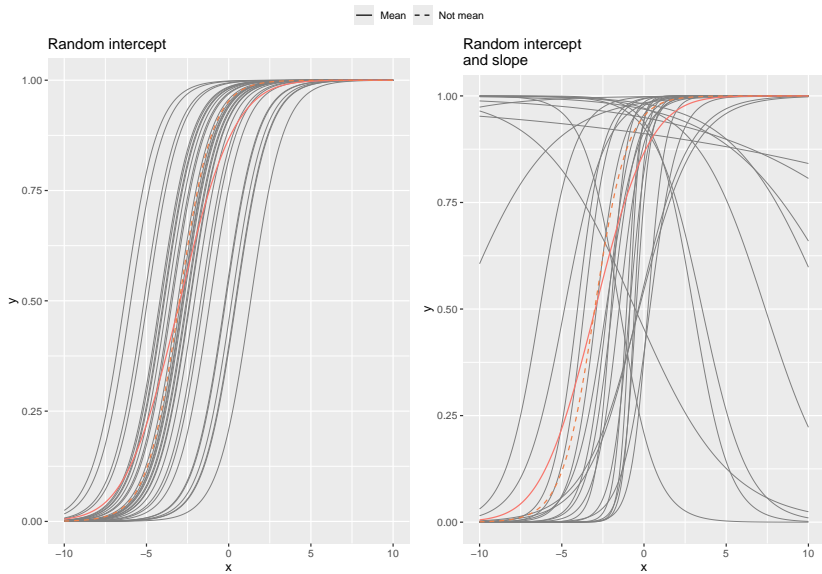
Comment on what you see.

Conditional and marginal model

To do parameter estimation in the LMM we started with the marginal model for Y_{ij}.

$$f(y_{ij}) = \int_{\gamma_i} f(y_{ij} \mid \gamma_i) f(\gamma_i) d\gamma_i$$

where

$$f(\gamma_i) \text{ is } N(\mathbf{0}, \mathbf{Q})$$

and $f(y_{ij} \mid \gamma_i)$ might be the binomial, Poisson, normal, gamma, … distribution - chosen as an exponential family.

It turns out that this integral can only be written out analytically in special cases.

In the special case that $f(y_{ij} \mid \gamma_i)$ is normal we saw in Module LMM that the marginal distribution is also multivariate normal.

Another special case is the log-linear Poisson random intercept where all $n_i = 1$. See pages 393-394 in our textbook to see the solution.

For the random intercept Poisson GLMM with log-link the expected marginal mean is (Agresti, 2015, page 309)

$$\mathsf{E}(Y_{ij}) = \mathsf{E}[\mathsf{E}(Y_{ij} \mid \gamma_{0i})] = \exp(\mathbf{x}_{ij}^T\beta + \tau_0^2/2)$$

The solid line in the plot to the left above gives this curve, while the dashed is $\exp(\mathbf{x}_{ij}^T\beta)$. In Agresti (2015, page 310) the variance of the marginal distribution is given as

$$\mathsf{Var}(Y_{ij}) = \cdots = \mathsf{E}(Y_{ij}) + [\mathsf{E}(Y_{ij})^2(\exp(\tau_0^2) - 1)]$$

so we see that the variance is larger than that of a Poisson distribution.

Remark: the conditional mean is $\exp(\mathbf{x}_{ij}^T\beta + \gamma_{0i})$, while the marginal mean is $\exp(\mathbf{x}_{ij}^T\beta + \tau_0^2/2)$.

(If the link function chosen had been the identity link then the marginal mean would have been $\mathbf{x}_{ij}^T\beta$.)

Parameter estimation

The most popular method for parameter estimation for regression is maximum likelihood. The likelihood of the data is given by the marginal distribution of all Y_{ij}s jointly. However, this can not be found analytically (as we saw above), so we must resort to numerical methods to evaluate the likelihood of a general GLMM.

The parameters we want to estimate are the fixed effects βs, and the parameters in $\mathbf{Q}$ for the random effects - denoted by ϑ.

The contribution from observations in the ith cluster is

$$f(\mathbf{y}_i \mid \beta, \vartheta) = \int_{\gamma_i} f(\mathbf{y}_i \mid \gamma_i, \beta) f(\gamma_i \mid \mathbf{Q}) d\gamma_i$$

$$= \int_{\gamma_i} \prod_{j=1}^{n_i} f(y_{ij} \mid \gamma_i, \beta) f(\gamma_i \mid \mathbf{Q}) d\gamma_i$$

And the likelihood is then (since the clusters are independent)

$$L(\beta, \vartheta) = \prod_{i=1}^{m} f(\mathbf{y_i} \mid \beta, \,)$$

To arrive at parameter estimates we need to maximize the likelihood with respect to β and ϑ, which is a complicated numerical problem, because wee need to integtrate over the γ_i's.

Let $l(\mathbf{y}_i, \gamma_i \mid \theta) = \log((\mathbf{y}_i, \gamma_i \mid \beta, \vartheta))$ be the *joint log-likelihood of the data and the random effects* (with $\theta = (\beta, \vartheta)$).

Observe that $l(\mathbf{y}_i, \gamma_i \mid \beta, \vartheta)$ depends on the unknown random effects γ_i and the parameters in $\theta = (\beta, \vartheta)$ (here we have the fixed effects and the parameters in $\mathbf{Q}$ for the random effects).

The maximum likelihood estimate for θ maximises

$$L(\theta) = \int_{\gamma} \exp(l(\mathbf{y}_i, \gamma_i \mid \theta))d\gamma$$

with respect to θ.

The random effects γ_i are now integrated out and the marginal likelihood $L(\theta)$ is the likelihood of the data as a function of the parameters only.

Approximations

There are a few approaches to approximating the integral:

▶ The Laplace Approximation
▶ Gauss-Hermite Quadrature
▶ Simulation (usually MCMC: see TMA4300 for more)
▶ INLA (Bayesian, much more advance. Involves nested Laplace approximations)

`glmer` function in `lme4` uses the Laplace approximation as a default, but can also use Gauss-Hermite Quadrature. Bayesian methods abound: MCMC is slow, but more flexible.

The Laplace Approximation

The Laplace Approximation assumes that $l(\mathbf{y}_i, \gamma \mid \theta)$ is quadratic in γ: this is equivalent to either

▶ assuming a Gaussian distribution for the sampling distribution of each γ_i, or

▶ using a single quadrature point in Gauss-Hermite Quadrature

The Approximation

Let $g(\gamma) = \ln(f(y \mid \gamma)f(\gamma))$, and look at the second order Taylor expansion around $\hat{u}(\theta)$:

$$g(\gamma_i) \approx \tilde{g}(\gamma_i) = g(\hat{\gamma}_i) + (\gamma_i - \hat{\gamma}_i)g'(\hat{\gamma}_i) + \frac{1}{2}(\gamma_i - \hat{\gamma}_i)^2 g''(\hat{\gamma}_i)$$

$$= g(\hat{\gamma}_i) - \frac{1}{2}(\gamma_i - \hat{\gamma}_i)^2 (-g''(\hat{\gamma}_i))$$

This means that $\exp(\tilde{g}(\gamma_i))$ is proportional to the normal density with mean $\mu = \hat{\gamma}_i$ and variance $\sigma^2 = -1/g''(\hat{\gamma}_i)$.
Putting this back into our integral:

$$L(\theta) = \int_\gamma \exp(l(\mathbf{y}_i, \gamma_i \mid \theta))d\gamma \approx \int_\gamma \exp(\tilde{l}(\mathbf{y}_i, \gamma_i \mid \theta))d\gamma$$

$$= \exp(l(\mathbf{y_i}, \hat{\gamma} \mid \theta)) \int_\gamma \exp(-\frac{1}{2}(\gamma - \hat{\gamma})^2)\mathbf{d}\gamma$$

$$= \exp(l(\mathbf{y}_i, \hat{\gamma} \mid \theta)) \sqrt{-2\pi/g''(\hat{\gamma}_i)}$$

Now we can plus this into our full likelihood, pausing only to note that $H(\theta)$ is the Hessian of $l(\mathbf{y}_i, \gamma_i \mid \theta)$ evaluated at $\hat{\gamma}(\theta)$, so contains $g''(\hat{\gamma}_i)$ terms

Let $\hat{\gamma}(\theta)$ be the maximum of $l(\mathbf{y}_i, \gamma_i \mid \theta)$ w.r.t γ.

The Laplace approximation for the marginal likelihood $L(\theta)$ is then

$$L^*(\theta) = (2\pi)^{n/2}\mathsf{det}(H(\theta))^{-1/2}\exp(f(\hat{u}, \theta))$$

Gauss-Hermite Quadrature

If the dimension of γ_i is small, for example our random intercept model, then the Gauss-Hermite quadrature may be used to approximate the integral above. This is because the form

$$\int_{-\infty}^{\infty} h(\gamma) \exp(-\gamma^2) d\gamma$$

can be written as a sum of weights and quadrature points that are roots of Hermite polynomials. This can then be maximized using Newton-Raphson methods.

When the dimension of γ gets large, this gets too slow, so other methods are usually better.

In essence, this is a better approximation than the Laplace Approximation, but there is a computational cost that might not be worth paying.

What have we not covered?

- We have skipped a lot of technicalities about the parameter estimation.
- How to predict the random effects γ_i: "the conditional modes" of the random effects.
- What is fitted values, and how are residuals calculated.
- How to test hypotheses?
- AIC to compare models (very similar to LMM).
- And, surely, much more.

Summing up - what have we learned about the GLMM?

▶ The GLMM can be formulated using three ingredients:
  ▶ distributional assumption: $f(y_{ij} \mid \gamma_i)$ from exponential family
  ▶ structural assumptions: linear predictor $\eta_{ij} = \mathbf{x}_{ij}^T \beta + \mathbf{u}_{ij} \gamma_i$ and link function $\eta_{ij} = g(\mu_{ij})$ where $\mu_{ij} = \mathsf{E}(Y_{ij} \mid \gamma_i)$.
  ▶ distributional assumptions for the random effects: $\gamma_i \sim N(\mathbf{0}, \mathbf{Q})$.
▶ The GLMM likelihood function is expressed as an integral with respect to the random effects and does (in general) not have a closed form solution.
▶ Numerical approximation methods need to be used to find parameter estimates, and one possibility is to use the Laplace approximation, but many competing method exists.

▶ Three R packages that can be investigated is `lme4`(function `glmer`) and `glmmTMB` (template model builder), and the NTNU-flagship `inla`. How to use these three packages on a simulated data set (binary data, logit link, random intercept and slope) is shown in the end of the module page (NOT on our reading list but for the interested student).

R packages

```
# install.packages('arm')
install.packages("reshape2")
install.packages("sp")
install.packages("glmmTBM")
install.packages("INLA", repos = "https://inla.r-inla-downl
install.packages("lme4")
install.packages("devtools")
library(devtools)
# install_github('romunov/AED')
install.packages("sjPlot")
install.packages("sjmisc")
```

Further reading

- Bolker et al. (2008):*Generalized linear mixed models: a practical guide for ecology and evolution*
- Zuur et al. (2009): "Mixed Effects Models and Extensions in Ecology with R", chapter 13 (pages 323-341). Available as free ebook from Springer for NTNU students. More explanations and less mathematics than Fahrmeir et al (2013), more focus on understanding. Link to ebook Chapter 13
- Agresti (2015): Foundations of linear and generalized linear models, Chapter 9.