Beyond GLMs: Mixed effects models

Bert van der Veen

Department of Mathematical Sciences, NTNU

Questions about yesterday?

A typical workflow

1) Collect data
2) Fit a model
3) Check assumptions
4) Perform inference
5) Write article

A typical workflow

1) Collect data
2) Fit a model
3) Check assumptions
4) Perform inference
5) Write article

Recap

- GLMs assume independence

Recap Likelihood: independence

$$\mathcal{L}(\mathbf{y};\Theta) = \prod_i^n f(y_i;\Theta) \tag{1}$$

We just multiply! (assumes independence)

What are random effects used for?

- An effect is "unobserved" or "latent"
- We do not want our model to "depend" on the effect
- To account for pseudoreplication
- Something that is nuisance (e.g., library depth when not measured)
- To incorporate correlation between observations (space, time, individuals, genes, ...)
- To estimate variation of an effect
- To incorporate overdispersion
- To estimate fixed effects at the "population level"

There is more.

Why is it called a random effect?

"Random" is in contract to "fixed"

- Fixed effects we maximize
- Random effects we marginalize

Random effects are random or stochastic; like our data they come from a distribution

Why is it called a random effect?

"Random" is in contract to "fixed"

- Fixed effects we maximize
- Random effects we marginalize

Random effects are random or stochastic; like our data they come from a distribution

Mixed effects means a model that incorporates both fixed and random effects

"Technicalities"

- We can formulate the same models
- But now, random effects are from a distribution
- This allows us to incorporate correlation

Our new likelihood

$$\mathcal{L}(\mathbf{y};\mathbf{\Sigma},\Theta) = \int \prod_{i}^{n} f(y_i;\Theta|\mathbf{u}) f(\mathbf{u};\mathbf{\Sigma}) d\mathbf{u} \qquad (2)$$

- Fixed effects: what we had so far
- Random effects: new, come from a distribution
- Mixed effects: contains both

Also notice that we cannot work with the log-likelihood here :(

The mixed effects model

$$\boxed{g}\left\{\boxed{\mathsf{E}(\mathbf{y}|\mathbf{u})}\right\} = \mathbf{X}\boxed{\beta} + \mathbf{Z}\boxed{\mathbf{u}} \tag{3}$$

1. Link-function
2. Conditional mean
3. Fixed effects parameter vector
4. Random effects parameter vector

The mixed effects model

$$g\{\mathrm{E}(\mathbf{y}|\mathbf{u})\} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} \tag{3}$$

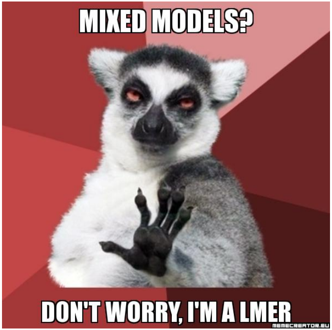
1. Link-function
2. Conditional mean
3. Fixed effects design matrix
4. Random effects design matrix

The random effects design matrix

- it’s the kind of thing as the fixed effects design matrix!

A linear mixed effects model

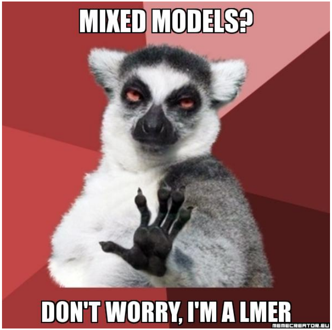
$$\boxed{\mathrm{E}(\mathbf{y}|\mathbf{u})} = \boldsymbol{\mu} \qquad (4)$$



A linear mixed effects model

$$\boxed{\mathrm{E}(\mathbf{y}|\mathbf{u})} = \boldsymbol{\mu} \tag{4}$$

$$\boxed{\mathbf{y}} = \boxed{\mathbf{X}\beta} + \boxed{\mathbf{Zu}} + \boxed{\mathbf{e}} \tag{5}$$

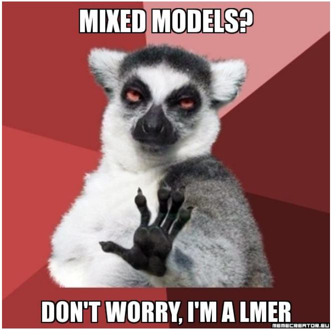


A linear mixed effects model

$$E(\mathbf{y}|\mathbf{u}) = \boldsymbol{\mu} \quad (4)$$

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{e} \quad (5)$$

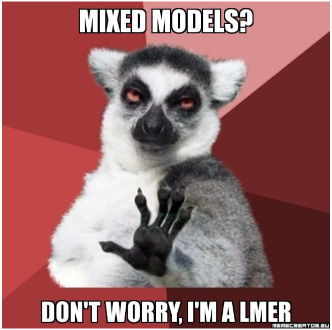
with $\mathbf{u} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$
with $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma^2)$



A linear mixed effects model

We can rewrite the model in terms of the complete error term.

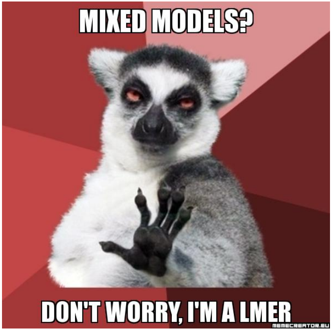
$$\mathbf{y} = \mathbf{X}\beta + \epsilon \qquad (6)$$



A linear mixed effects model

We can rewrite the model in terms of the complete error term.

$$\boxed{\mathbf{y}} = \boxed{\mathbf{X}\beta} + \underbrace{\boxed{\epsilon}}_{\mathcal{N}(0,\, \mathbf{Z}\Sigma\mathbf{Z}^\top + \mathbf{I}\sigma^2)} \qquad (6)$$



So, we are including covariance between our errors in the model.

The objective function

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Zu} + \mathbf{e} \tag{7}$$

with $\epsilon = \mathbf{Zu} + \mathbf{e}$ and $\epsilon \sim \mathcal{N}(0, \mathbf{Z\Sigma Z}^\top + \mathbf{I}\sigma^2)$

we have the marginal distribution $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\beta, \mathbf{Z\Sigma Z}^\top + \mathbf{I}\sigma^2)$

This is not how things are done in practice (because the covariance matrix can get quite big!)

Estimation

We assume that the random effects follow a normal distribution:

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

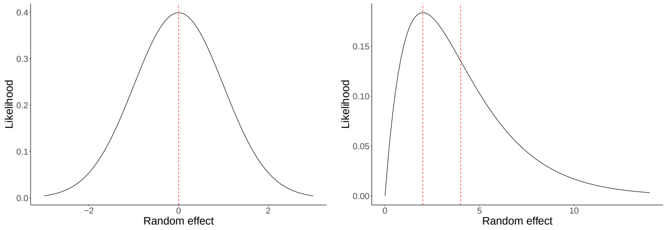
However, we are really interested in $u|\mathbf{y}$ instead; this holds information from our data!

- We want $f(\mathbf{u}|\mathbf{y})$ to be as flexible as possible
- So we can capture the patterns in the data as accurately as possible

Estimation methods

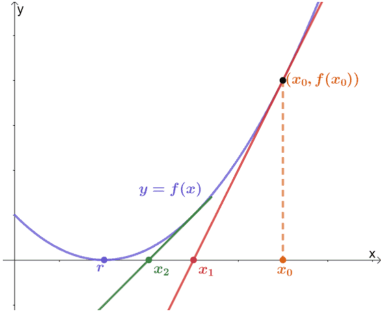
- Penalized quasi-likelihood methods
- Adaptive GH quadrature
- Laplace approximation
- Variational approximations
- Et cetera (see e.g., Bolker et al. 2009)

Measure of central tendency: Mean or Mode



Fitting

- Objective function is assumed to be quadratic (often)



- Usually navigate it with newton's method (numerical optimization)

Recall: Maximum Likelihood Estimation

At the maximum of the likelihood:

- The gradient is zero (tangent is straight)
- The hessian (of -LL) should
  - have positive diagonals
  - positive eigenvalues
  - be symmetric
  - and is thus invertible (we go up in both directions)
- Asymptotic covariance matrix is given by the inverse of the negative Hessian

These are important concepts to understand error messages and convergence in mixed effects models.

There are many R-packages

- `nlme`
- `lme4`
- `glmmTMB` (or `glmmADMB`)
- `sdmTMB`
- `MASS`
- `glmmML`
- `repeated`
- `glmm`
- `hglm`
- `spaMM`
- `gllvm`
- `mcmcGLMM`
- `INLA`
- `inlabru`
- MCMC frameworks (JAGS, STAN, NIMBLE, greta)

`lme4` and `glmmTMB` are most commonly used.

`lme4` (Bates et al. 2015)

- Correlation between random effects
- Sparse matrices
- Modern matrix algebra libraries
- Likelihood profiling

But can be fussy about convergence

glmmTMB (Brooks et al. 2017)

- Correlation between and within random effects (e.g., spatial)
- Uses state-of-the art AD software (TMB, Kristensen et al. 2015)
- More supported distribution
  - E.g. zero-inflation
  - Double hierarchical GLMs

Specification with formula syntax in R

- We can think of our model in the same way
  - Intercepts for categorical covariates
  - Slopes for continuous covariates
  - Interactions
- Now the "parameters" can be correlated
- With the R syntax we formulate:
  - The design matrix $\mathbf{Z}$
  - The covariance matrix $\boldsymbol{\Sigma}$
- Just as before: intercepts are categorical, slopes for continuous covariates

Random effects R formula

Now some examples of how it works in R. Generally:

`y ~ (continuous and/or categorical | categorical)`

"Nested":

`y ~ (1|a/b)` is the same as `y ~ (1|a:b + b)`

"Crossed":

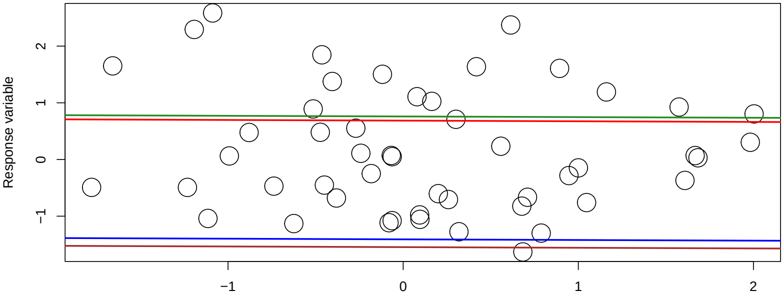
`y ~ (1|a) + (1|b)`

A few examples next

Random intercepts

$$y_{ij} = \mathbf{x}_i \boldsymbol{\beta} + \alpha_j, \qquad \text{with } \alpha_j \sim \mathcal{N}(0, \sigma^2)$$

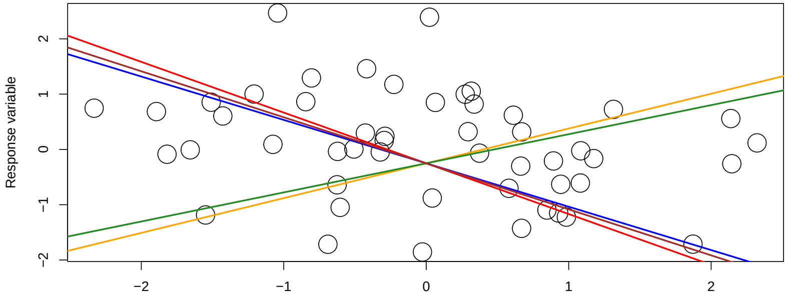
```
y ~ fixed effects + (1|random intercept)
```



Random slopes

$$y_{ij} = \mathbf{x}_i\boldsymbol{\beta} + z_i u_j, \qquad \text{with } u_i \sim \mathcal{N}(0, \sigma^2)$$

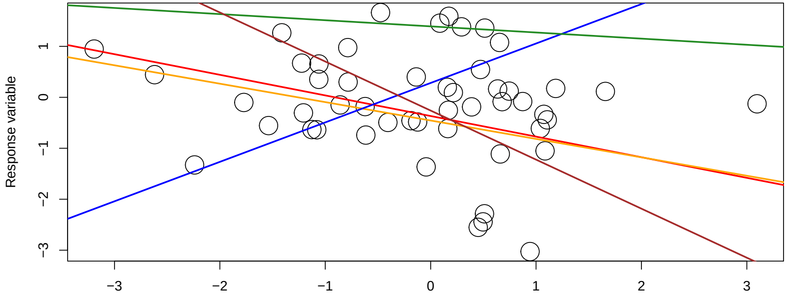
```
y ~ fixed effects + (0+random slope|categories)
```



Random intercepts and slopes

$$y_{ij} = \mathbf{x}_i\beta + \alpha_j + z_i u_j, \text{ with } \begin{pmatrix} \alpha_j \\ u_j \end{pmatrix} \sim \mathcal{N}\left\{\mathbf{0}, \begin{pmatrix} \sigma_1^2 & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \end{pmatrix}\right\}$$

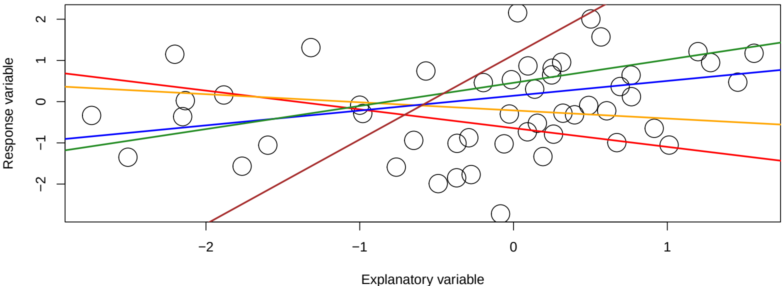
```
y ~ fixed effects + (1|random intercept)+(0+random slope|categories)
```



Correlated random intercepts and slopes

$$y_{ij} = \mathbf{x}_i \beta + \alpha_j + z_i u_j, \text{ with } \begin{pmatrix} \alpha_j \\ u_j \end{pmatrix} \sim \mathcal{N}\left\{ \mathbf{0}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \right\}$$

```
y ~ fixed effects + (random slope|random intercept)
```



Example: Owls data in `glmmTMB`

Originally by Roulin and Bersier (2007)

- Count of begging attempts by chicks
- Also data on treatments, nest ID, sex of the parent, and broodsize

Nest	FoodTreatment	SexParent	ArrivalTime	SiblingNegotiation	Br
AutavauxTV	Deprived	Male	22.25	4	
AutavauxTV	Satiated	Male	22.38	0	
AutavauxTV	Deprived	Male	22.53	2	
AutavauxTV	Deprived	Male	22.56	2	
AutavauxTV	Deprived	Male	22.61	2	
AutavauxTV	Deprived	Male	22.65	2	
AutavauxTV	Deprived	Male	22.76	18	
AutavauxTV	Satiated	Female	22.90	4	
AutavauxTV	Deprived	Male	22.98	18	
AutavauxTV	Satiated	Female	23.07	0	

Example: random intercept in glmmTMB

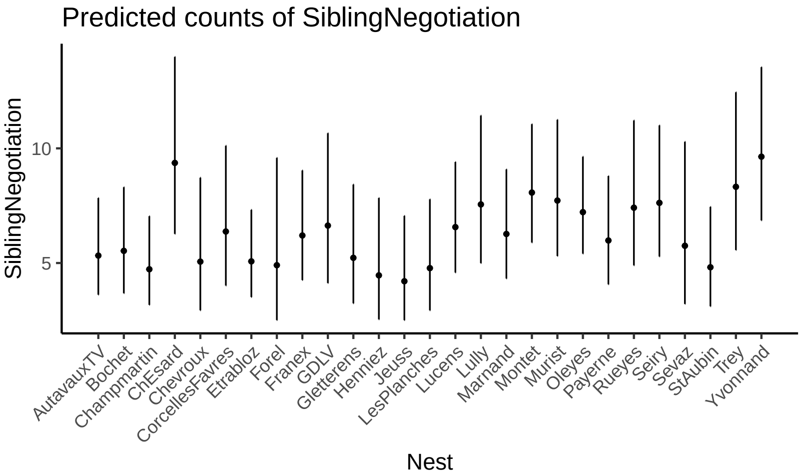
```
##  Family: nbinom2  ( log )
## Formula:          SiblingNegotiation ~ (1 | Nest)
## Data: Owls
##
##      AIC      BIC   logLik deviance df.resid
##   3533.1   3546.3  -1763.5   3527.1      596
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev.
##  Nest   (Intercept) 0.09591  0.3097
## Number of obs: 599, groups:  Nest, 27
##
## Dispersion parameter for nbinom2 family (): 0.75
##
## Conditional model:
##             Estimate Std. Error z value Pr(>|z|)
```

Example: extract random effect in glmmTMB

```r
ranef(model1)
```

```
## $Nest
##                   (Intercept)
## AutavauxTV       -0.154702417
## Bochet           -0.116696108
## Champmartin      -0.273447716
## ChEsard           0.410074693
## Chevroux         -0.205451776
## CorcellesFavres   0.025333070
## Etrabloz         -0.203155670
## Forel            -0.236717059
## Franex           -0.002254524
## GDLV              0.065028688
## Gletterens       -0.173183282
## Henniez          -0.331197524
## Jeuss            -0.390415694
```

Example: plot random effect in `glmmTMB`



Combining fixed and random effects

Having the same covariate in the fixed and random effects, is equal to having a mean effect, and variation around that

On the previous example: global intercept with nest-specific deviation

We can do this for slopes too (not covered in these examples)

Example: random slopes in glmmTMB

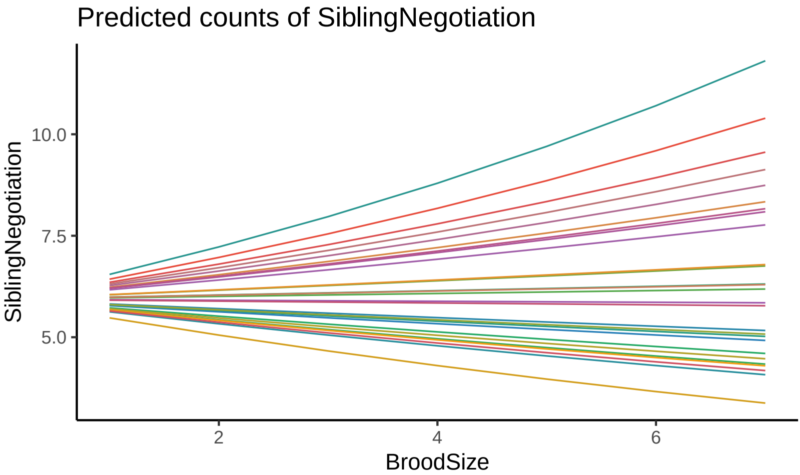
```
##  Family: nbinom2  ( log )
## Formula:          SiblingNegotiation ~ (0 + BroodSize | Nest)
## Data: Owls
##
##      AIC      BIC   logLik deviance df.resid
##   3533.2   3546.4  -1763.6   3527.2      596
##
## Random effects:
##
## Conditional model:
##  Groups Name      Variance Std.Dev.
##  Nest   BroodSize 0.004544 0.06741
## Number of obs: 599, groups:  Nest, 27
##
## Dispersion parameter for nbinom2 family (): 0.746
##
## Conditional model:
##            Estimate Std. Error z value Pr(>|z|)
```

Example: extract random effect in glmmTMB

```
ranef(model2)
```

```
## $Nest
##                  BroodSize
## AutavauxTV     -0.026764924
## Bochet         -0.019826801
## Champmartin    -0.053656369
## ChEsard         0.098280324
## Chevroux       -0.023913350
## CorcellesFavres 0.008748365
## Etrabloz       -0.036398867
## Forel          -0.044787068
## Franex          0.005923827
## GDLV            0.018469824
```

Example: plot it with `ggeffects`



Example: random slopes in glmmTMB

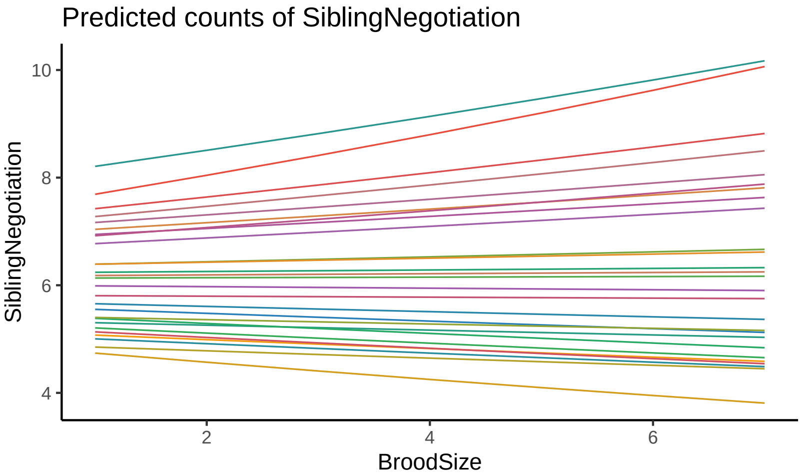
```
##  Family: nbinom2  ( log )
## Formula:          SiblingNegotiation ~ (BroodSize || Nest)
## Data: Owls
##
##      AIC      BIC   logLik deviance df.resid
##   3534.8   3552.4  -1763.4   3526.8      595
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev. Corr
##  Nest   (Intercept) 0.055801 0.2362
##         BroodSize   0.001893 0.0435   0.00
## Number of obs: 599, groups:  Nest, 27
##
## Dispersion parameter for nbinom2 family (): 0.749
##
## Conditional model:
```

Example: extract random effect in glmmTMB

```
ranef(model3)
```

```
## $Nest
##                  (Intercept)    BroodSize
## AutavauxTV      -0.078914423 -0.0133831926
## Bochet          -0.064845645 -0.0087978012
## Champmartin     -0.178163213 -0.0181289491
## ChEsard          0.263158987  0.0357035612
## Chevroux        -0.129255881 -0.0087682646
## CorcellesFavres  0.022424439  0.0022817927
## Etrabloz        -0.105057944 -0.0178169040
## Forel           -0.137744812 -0.0186882477
## Franex           0.006561821  0.0008902617
## GDLV             0.041464534  0.0070320206
## Gletterens      -0.112220377 -0.0076126359
## Henniez         -0.212817836 -0.0144368139
## Jeuss           -0.214329310 -0.0363483675
```

Example: plot it with `ggeffects`



Example: random slopes and random intercepts in glmmTMB

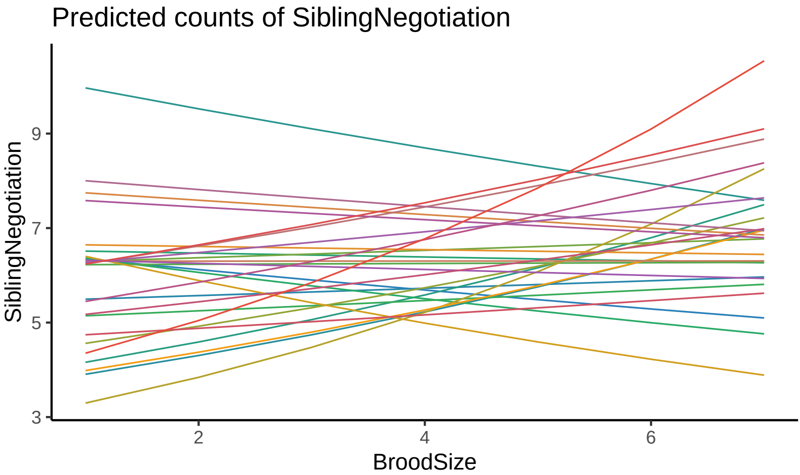
```
##  Family: nbinom2  ( log )
## Formula:          SiblingNegotiation ~ (BroodSize | Nest)
## Data: Owls
##
##      AIC      BIC   logLik deviance df.resid
##   3536.0   3557.9  -1763.0   3526.0      594
##
## Random effects:
##
## Conditional model:
##  Groups Name        Variance Std.Dev. Corr
##  Nest   (Intercept) 0.52815  0.7267
##         BroodSize   0.02541  0.1594   -0.95
## Number of obs: 599, groups:  Nest, 27
##
## Dispersion parameter for nbinom2 family (): 0.748
##
```

Example: extract random effect in glmmTMB

```
ranef(model4)
```

```
## $Nest
##                  (Intercept)     BroodSize
## AutavauxTV        0.043009382 -0.0365814235
## Bochet           -0.151705211  0.0137061099
## Champmartin      -0.576358671  0.0969512834
## ChEsard           0.502738309 -0.0454208954
## Chevroux         -0.514525480  0.0981411962
## CorcellesFavres   0.038313328 -0.0064448173
## Etrabloz          0.056787822 -0.0483006101
## Forel            -0.224348318  0.0202691963
## Franex           -0.014459691  0.0013063896
## GDLV             -0.014317234  0.0121774546
```

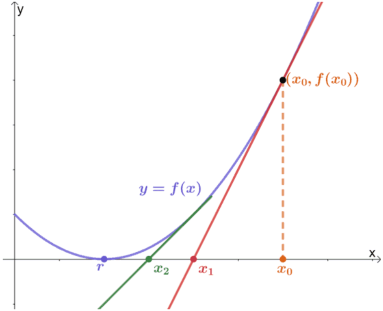
Example: plot it with `ggeffects`



Convergence

see Ben Bolker's GLMM FAQ, `lme4` page on performance, and the `glmmTMB` troubleshooting vignette

Assessing arrival at the MLE



Assessing arrival at the MLE

1. Stopping criteria
   - Maximum iterations
   - Gradient close to zero
   - Relative criterion: objective function value improvement
   - Absolute criterion: objective function becomes zero (say)
2. Gradient
3. Hessian

`lme4` warnings: hessian

- `Warning: Problem with Hessian check (infinite or missing values?)`
- `Warning: Hessian is numerically singular: parameters are not uniquely determined`
- `Warning: Model failed to converge: degenerate Hessian with 2 negative eigenvalues`
- `Warning: Model is nearly unidentifiable: very large eigenvalue - Rescale variables?`
- `Warning: Model is nearly unidentifiable: very large eigenvalue ratio - Rescale variables?`

Singular matrix

- determinant is zero
  - has zero eigenvalue(s)
- does not have inverse

$$\mathbf{HA} = \mathbf{I} \tag{8}$$

Numerical optimisation: best practices

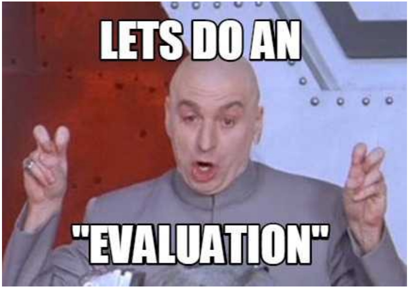
1. Standardise (center and scale) explanatory variables
2. Try different optimisation routines
3. Different starting values
4. Rethink your model

Mixed effects model troubleshooting

see Ben Bolker's GLMM FAQ

- check data for mistakes
- check model formulation
  - correct distribution and link-function
  - few random effects levels
  - few (non-zero) observations in a category
  - overly complex: drop terms with zero variances
- double-check hessian calculation (finite differences)
- use random effect as fixed effect
- '?convergence' (and see the last line "convergence issues" for large datasets)

Assumption checking



Residuals

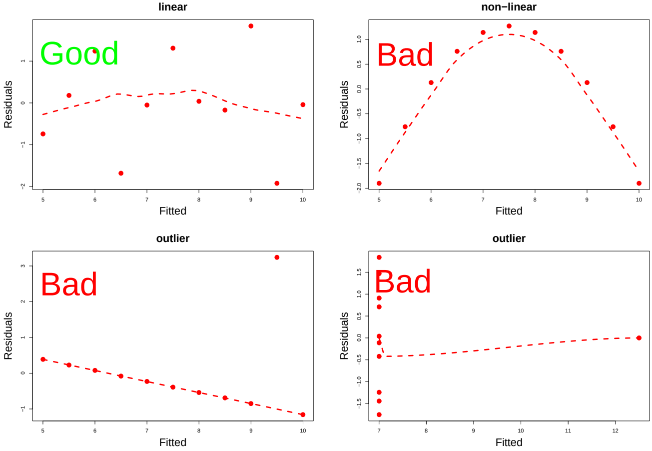
$$\mathbf{y} = \mathbf{X} + \mathbf{Zu} + \underbrace{\mathbf{e}}_{\text{Residual}} \qquad (9)$$

Residuals

$$\boxed{\mathbf{y}} = \boxed{\mathbf{X}} + \boxed{\mathbf{Zu}} + \underbrace{\boxed{\mathbf{e}}}_{\text{Residual}} \qquad (9)$$

Violated residual assumptions mean that some or all of your model's results are untrustworthy.

Residual diagnostics: residuals vs. fitted



GLMM residuals

Conditional

$$\mathsf{g}\{\mathsf{E}(y_{ij}|x_i)\} = \hat{\alpha} + \hat{\beta}x_i + z_i\hat{u}_j \tag{10}$$

Unconditional

$$\mathsf{g}\{\mathsf{E}(y_{ij}|x_i)\} = \hat{\alpha} + \hat{\beta}x_i \tag{11}$$

How do we calculate the residual?

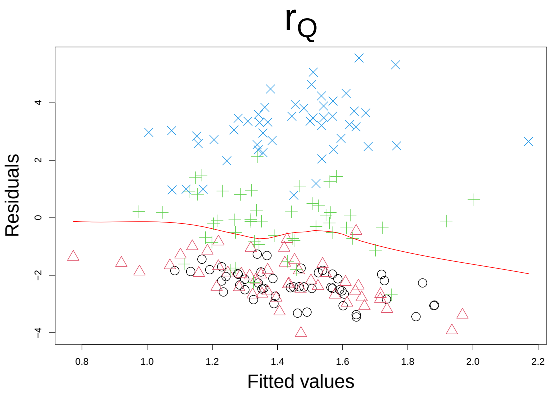
- should we condition on the predicted random effect?
- simulate from conditional distribution?

i.e. a range of options

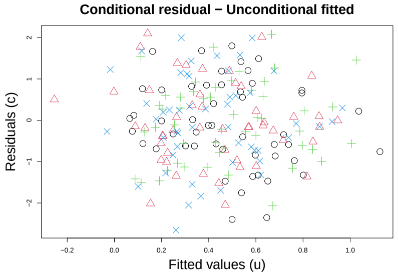
Simulation: grouping of errors

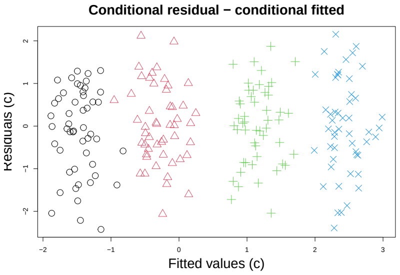
```
n <- 200
ngroups <- 4
alpha <- 0.5
beta <- -1
x <- rnorm(n, sd = 0.2)

fac<-rep(1:ngroups,each=n/ngroups)
e <- seq(from=-2,to=2,length.out=ngroups)
mu <- exp(alpha + beta*x + e[fac])
y <- rpois(n = n, lambda = mu)
```

Example: Poisson residuals (grouping)



Residual diagnostics: Poisson residuals (grouping)





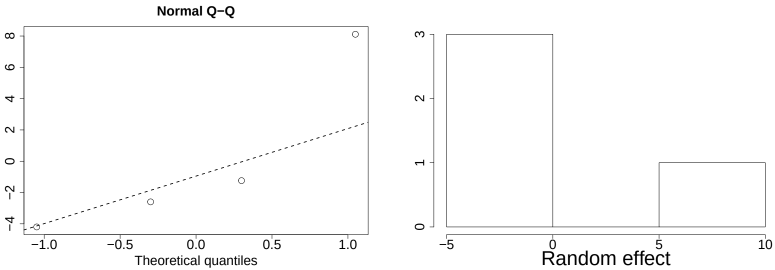
GLMM: checking random effect assumptions

- random effect is a type of residual
- $\hat{u}_j$ is an estimate of the mean or mode of $p(u_j|y_i)$
- we treat $\hat{u}_j$ as a sample of the random effect distribution
- so we check assumptions (marginal normality, constant variance, independence, no outliers)!
- difficult with small number of groups
- needs to be done for every random effect

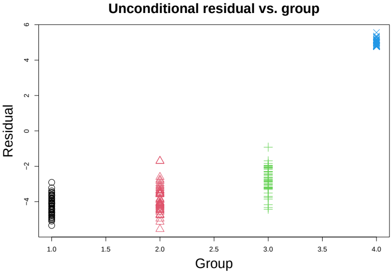
Simulation: GLMM (outlier)

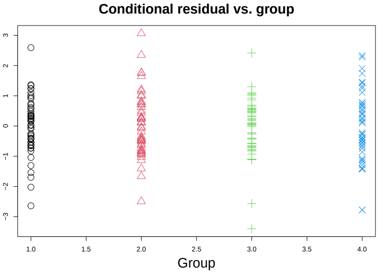
```
n <- 200
ngroups <- 4
alpha <- 0.5
beta <- -1
x <- rnorm(n, sd = 0.2)

fac<-rep(1:ngroups,each=n/ngroups)
e <- seq(from=-2,to=2,length.out=ngroups)
e[4] <- 10
mu <- exp(alpha + beta*x + e[fac])
y <- rpois(n = n, lambda = mu)
```

GLMM diagnostics



GLMM diagnostics

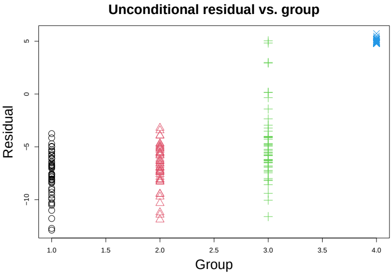


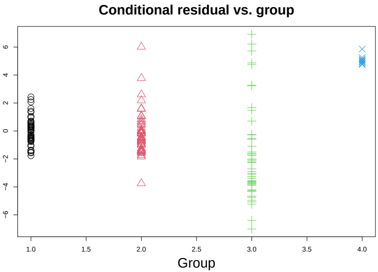


What if constant variance is violated

```
n <- 200
ngroups <- 4
alpha <- 0.5
beta <- -1
x <- rnorm(n, sd = 0.2)

fac<-rep(1:ngroups,each=n/ngroups)
e <- seq(from=-2,to=2,length.out=ngroups)
e[4] <- 10
e2 <- MASS::mvrnorm(1,rep(0,n), diag(rep(c(1,2,3,4),
                                        each=n/ngroups)))
mu <- exp(alpha + beta*x + e[fac] + e2)
y <- rpois(n = n, lambda = mu)
```

GLMM diagnostics





Residual checking for mixed effects models

- Check assumptions
  - Use both conditional and marginal residuals
  - Have a look at the `DHARMa` vignette
- Correct violations

Violation of some assumptions might be OK

Inference

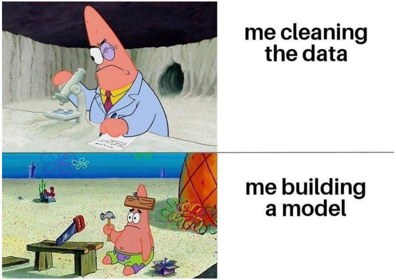
We have a good model!

Inference

We have a good model!

Now we want to do inference

- Hypothesis tests (t-test, LRT)/ P-values
- Model-selection (e.g., with AIC; Akaike 1973)
- Et cetera (R^2).



Next crisis

1. Estimate
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta}}{\sqrt{\sigma^2}} \tag{12}$$

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)}{\sqrt{\sigma^2}} \tag{12}$$

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)}{\sqrt{\hat{\sigma}^2}} \tag{12}$$

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)}{\sqrt{\hat{\sigma}^2 \sim \sigma^2 \chi^2_{n-r}}} \tag{12}$$

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)}{\sqrt{\hat{\sigma}^2 \sim \sigma^2 \chi^2_{n-r}}} \sim t_{n-r} \tag{12}$$

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top \mathbf{X})^{-1}\sigma^2)}{\sqrt{\hat{\sigma}^2 \sim \sigma^2 \chi^2_{n-r}}} \sim t_{n-r} \tag{12}$$

- P-value (probability for the absolute value of the statistic to take on this, or a more extreme, value)

Next crisis

1. Estimate $\sim \mathcal{N}(\beta, (\mathbf{X}^\top\mathbf{X})^{-1}\sigma^2)$
2. Truth

$$t = \frac{\boxed{\hat{\beta}} - \boxed{\beta} \sim \mathcal{N}(0, \mathbf{X}^\top\mathbf{X})^{-1}\sigma^2)}{\sqrt{\hat{\sigma}^2 \sim \sigma^2\chi^2_{n-r}}} \sim t_{n-r} \tag{12}$$

- P-value (probability for the absolute value of the statistic to take on this, or a more extreme, value)
- Degrees of freedom concept in (G)LMMs is unclear and subject to ongoing debate
- Unless in the simplest (balanced and uncorrelated) case

DF approximation

- as implemented in e.g., the `lmerTest` R-package

$$\hat{\sigma^2} = \frac{RSS}{n-r} \tag{13}$$

- i.i.f. $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, $\quad RSS/\sigma \sim \chi^2_{n-r}$
  - $F = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}/n1}{\boldsymbol{\epsilon}_2^\top \boldsymbol{\epsilon}_2/n2} \sim F_{n1,n2}$
- Otherwise approximate df
  - Satterthwaite (1941)
  - Kenward-Roger (1997)
    - REML only
- Performance is questionable
- Can always resort to asymptotics (Wald tests $\sqrt{z} = \frac{\hat{\beta}-\beta}{\sigma} \sim \mathcal{N}(0,1)$)

DF approximation

- as implemented in e.g., the `lmerTest` R-package

$$\hat{\sigma^2} = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}}{n - r} \tag{13}$$

- i.i.f. $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$, $\quad RSS/\sigma \sim \chi^2_{n-r}$
  - $F = \frac{\boldsymbol{\epsilon}^\top \boldsymbol{\epsilon}/n1}{\boldsymbol{\epsilon}_2^\top \boldsymbol{\epsilon}_2/n2} \sim F_{n1,n2}$
- Otherwise approximate df
  - Satterthwaite (1941)
  - Kenward-Roger (1997)
    - REML only
- Performance is questionable
- Can always resort to asymptotics (Wald tests $\sqrt{z} = \frac{\hat{\beta} - \beta}{\sigma} \sim \mathcal{N}(0, 1)$)

Information criteria

- Information criteria do not work "on the boundary"
- Not clear which measure of AIC works best (it depends)
- As usual beware model selection bias (Freedman's paradox)
- Generally proceed with caution

Take away tips

No free lunch in statistics

Take away tips

No free lunch in statistics

- Scale your predictors
- Carefully consider the model structure
- Keep your model as simple as possible, but not simpler
- Different packages have different benefits
  - `glmmTMB` vs. `lme4`
- Try not to -blindly- assume approximations perform well
- Always check (residual) assumptions
- Be pragmatic
- Consult a statistician

Write article?

- Be pragmatic (but not too much)
- Do not rely on P-values (they're mostly useless anyway)
- Report a single (full) model (if possible, see convergence)
  - Alternatively, describe your modeling procedure
  - Use information criteria e.g., to compare two competing (non-nested) hypotheses
- Nakagawa R^2 for mixed models (also approximation)

Focus on estimates (effect sizes) and statistical uncertainy

End