Generalized Linear Mixed Models

Analysis of Ecological and Environmental Data

QERM 514

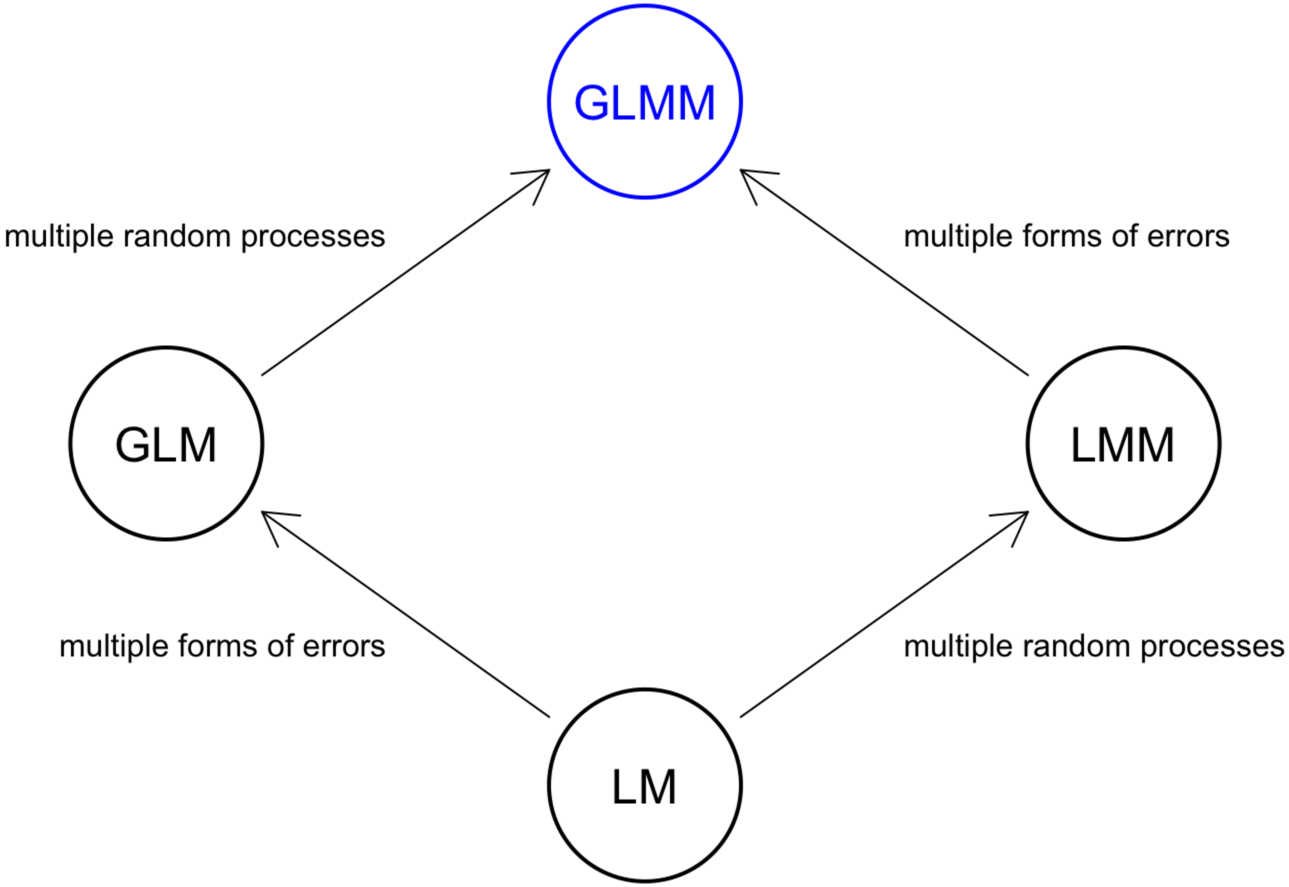
Mark Scheuerell

22 May 2020

Goals for today

- Understand the structural components of generalized linear mixed models
- Understand the options for fitting GLMMs and their pros and cons
- Understand some of the diagnostics available for evaluating GLMM fits

Forms of linear models



Generalized linear mixed model

GLMMs combine the flexibility of non-normal distributions (GLMs) with the ability to address correlations among observations and nested data structures (LMMs)

Generalized linear mixed model

Good news

- these extensions follow similar methods to GLMs and LMMs

Bad news

- these models are on the frontier of statistical research
- existing documentation is rather technical
- multiple approaches for fitting models; some with different results

Generalized linear mixed model

Just like GLMs, GLMMs have three components:

1. Distribution of the data $f(y;\theta)$
2. Link function $g(\eta)$
3. Linear predictor η

Linear predictor for a GLM

We can write the linear predictor for GLMs as

$$
\eta = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \\
\Downarrow \\
\eta = \mathbf{X}\boldsymbol{\beta}
$$

where the β_i are fixed effects of the covariates x_i

Linear predictor for a GLMM

For GLMMs, our linear predictor also includes random effects

$$\eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \alpha_0 + \alpha_1 z_1 + \cdots + \alpha_l z_l$$
$$\Downarrow$$
$$\eta = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha}$$

where the β_i are fixed effects of the covariates x_i

Generalized linear mixed model

Survival of fish $s_{i,j}$ as a function of length $x_{i,j}$ in some location j

$$\begin{aligned}
\text{data distribution:}\quad & y_{i,j} \sim \text{Binomial}(N_{i,j}, s_{i,j}) \\
\text{link function:}\quad & \text{logit}(s_{i,j}) = \log\left(\frac{s_{i,j}}{1 - s_{i,j}}\right) = \mu_{i,j} \\
\text{linear model:}\quad & \mu_{i,j} = (\beta_0 + \alpha_j) + \beta_1 x_{i,j} \\
& \alpha_j \sim \text{N}(0, \sigma_\delta^2)
\end{aligned}$$

Generalized linear mixed model

Best practices suggest we try to keep things simple

Why? Because GLMMs involve solving an integral with no analytical solution

Likelihood for GLMMs

Recall that we think of likelihoods in terms of the *observed data*

But the random effects in our model are *unobserved* random variables, so we need to integrate them out of the likelihood

Likelihood for GLMMs

The likelihood for a GLMM involves integrating over all possible random effects

$$\mathcal{L}(y; \boldsymbol{\beta}, \phi, \boldsymbol{\nu}) = \prod_i \int \underbrace{f_d(y; \boldsymbol{\beta}, \phi, \boldsymbol{\alpha})}_{\text{distn for data}} \; \underbrace{f_r(\boldsymbol{\alpha}; \boldsymbol{\nu})}_{\text{distn for RE}} \, d\boldsymbol{\alpha}$$

If $f(y; \boldsymbol{\beta}, \phi, \boldsymbol{\alpha})$ is not Gaussian, we cannot remove it from the likelihood, which makes it *very* difficult to compute

Approaches to fitting GLMMs

To avoid the integral, we will consider 3 methods that approximate the likelihood

They all have pros and cons so it’s not possible to pick the “best”

Penalized quasi-likelihood

Penalized quasi-likelihood (PQL) uses a Taylor series expansion to approximate the linear predictor as an LMM

$$\begin{aligned} g(\boldsymbol{\mu}) &= \boldsymbol{\eta} \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} \\ &\Downarrow \\ g(\mathbf{y}) &\approx g(\boldsymbol{\mu}) + g'(\boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu}) \\ &\approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + g'(\boldsymbol{\mu})\boldsymbol{\epsilon} \end{aligned}$$

Penalized quasi-likelihood

The conditional variance of the data in a GLMM is then

$$
\begin{aligned}
g(\mathbf{y}) &\approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + g'(\boldsymbol{\mu})\boldsymbol{\epsilon} \\
&\Downarrow \\
g(\mathbf{y}) - \mathbf{X}\boldsymbol{\beta} &\approx \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} g'(\boldsymbol{\mu}) \\
&\Downarrow \\
\text{Var}\left(g(\mathbf{y}) - \mathbf{X}\boldsymbol{\beta}\right) &\approx \text{Var}\left(\mathbf{Z}\boldsymbol{\alpha}\right) + \text{Var}\left(\boldsymbol{\epsilon} g'(\boldsymbol{\mu})\right)
\end{aligned}
$$

Penalized quasi-likelihood

The conditional variance of the data in a GLMM is then

$$
\begin{gathered}
g(\mathbf{y}) \approx \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + g'(\boldsymbol{\mu})\boldsymbol{\epsilon} \\
\Downarrow \\
g(\mathbf{y}) - \mathbf{X}\boldsymbol{\beta} \approx \mathbf{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon} g'(\boldsymbol{\mu}) \\
\Downarrow \\
\text{Var}\left(g(\mathbf{y}) - \mathbf{X}\boldsymbol{\beta}\right) \approx \text{Var}\left(\mathbf{Z}\boldsymbol{\alpha}\right) + \text{Var}\left(\boldsymbol{\epsilon} g'(\boldsymbol{\mu})\right)
\end{gathered}
$$

which is similar to that for an LMM

$$\text{Var}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right) = \text{Var}\left(\mathbf{Z}\boldsymbol{\alpha}\right) + \text{Var}\left(\boldsymbol{\epsilon}\right)$$

Penalized quasi-likelihood

Pros

- fast, flexible, and widely implemented

Cons

- only asymptotically correct
- biased for Binomial and Poisson with small samples
- inference confounded by approximate likelihood

Laplace approximation

Laplace approximation is a long standing (1774) method for computing integrals of the form

$$\int f(x)e^{\lambda g(x)}dx$$

This integrand is quite similar to the likelihood of a GLMM based on exponential distributions

Thus, we only need to find the maximum of $g(x)$ and its second derivative, and evaluate them at only one point

Laplace approximation

Pros

- approximation of true likelihood rather than quasi-likelihood
- more accurate than PQL

Cons

- slower and less flexible than PQL
- may be impossible to compute for complex models

Gauss-Hermite quadrature

Gauss-Hermite quadrature is an expansion of Laplace approximation where the integrand is evaluated at more than one point

Quadrature is a method for numerically approximating an integral as a weighted sum

$$\int f(u)e^{-u^2}\,du \approx \sum_i w_i f(u_i)$$

This method works by optimizing the placement and number of the u_i and the choice of the w_i

Gauss-Hermite quadrature

Pros

- More accurate than Laplace

Cons

- Slow and computationally intense
- Limited to a few random effects (one in practice)

Fitting GLMMs

Example

Let’s consider a long-term study of invasive brown tree snakes in Guam

Introduced to the island shortly after WWII

Voracious predators on native birds and other vertebrates

Photo by Pavel Kirillov

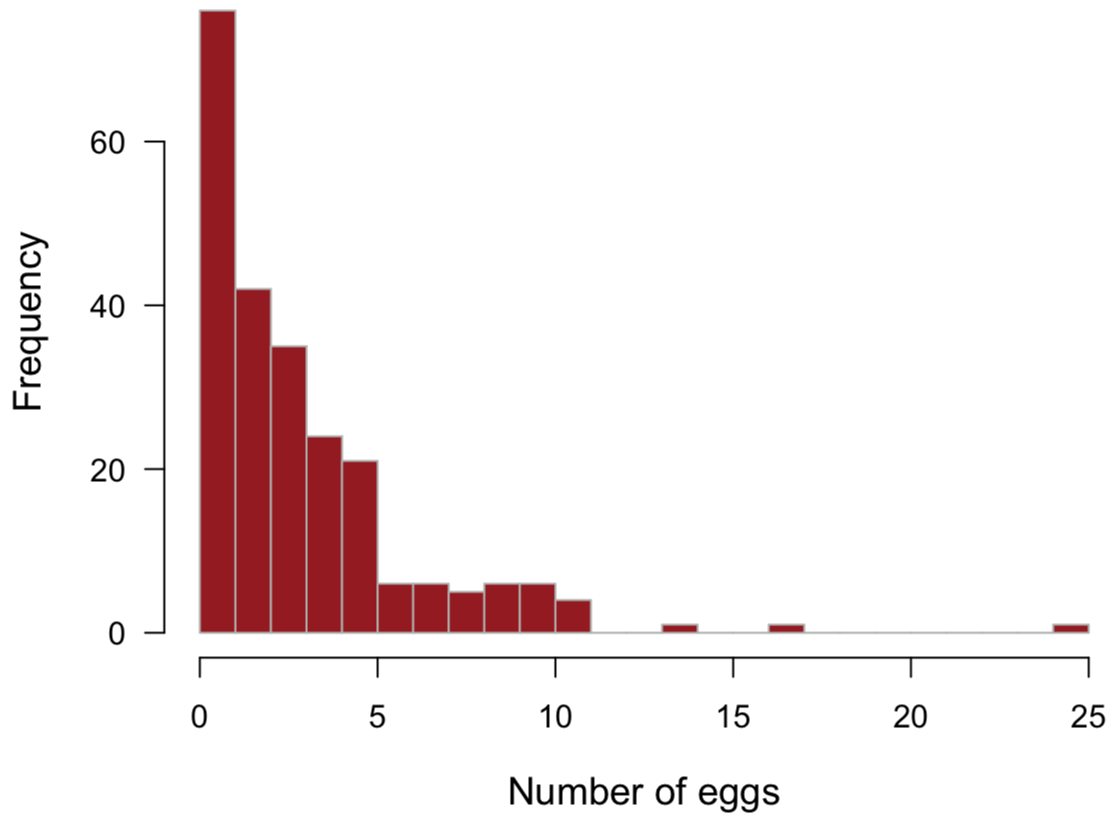
Brown tree snakes

Our data consist of counts of the number of eggs per female at 23 locations over 14 years

We are interested in the fixed effect of body size and the random effects of location and year

We'll begin with only the effects of body size and location

Brown tree snakes



Brown tree snakes

Penalized quasi-likelihood

We fit PQL models with `MASS::glmPQL()`

```
## load MASS
library(MASS)
## fit model
snakes_pql <- glmmPQL(eggs ~ size, random = ~1 | loc, data = df_eggs,
                      family = poisson)
```

```
summary(snakes_pql)
```

Brown tree snakes

```
## Linear mixed-effects model fit by maximum likelihood
##  Data: df_eggs
##   AIC BIC logLik
##    NA  NA     NA
##
## Random effects:
##  Formula: ~1 | loc
##         (Intercept) Residual
## StdDev:   0.5077229 1.183238
##
## Variance function:
##  Structure: fixed weights
##  Formula: ~invwt
## Fixed effects: eggs ~ size
##                 Value  Std.Error  DF  t-value p-value
## (Intercept) 1.1247363 0.11687536 210 9.623383       0
## size        0.5079533 0.07916825 210 6.416124       0
##  Correlation:
##      (Intr)
## size -0.069
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -1.7744344 -0.7176552 -0.2481373  0.5028263  3.3994803
##
## Number of Observations: 234
## Number of Groups: 23
```

Brown tree snakes

Laplace

We can fit Laplace models with `lme4::glmer()`

```
## load lme4
library(lme4)
## fit model
snakes_lap <- glmer(eggs ~ size + (1 | loc), data = df_eggs,
                    family = poisson)
```

```
summary(snakes_lap)
```

Brown tree snakes

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: eggs ~ size + (1 | loc)
##    Data: df_eggs
##
##      AIC      BIC   logLik deviance df.resid
##   1006.7   1017.1   -500.4   1000.7      231
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1158 -0.8480 -0.2741  0.5931  4.0679
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  loc    (Intercept) 0.2753   0.5247
## Number of obs: 234, groups:  loc, 23
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.09929    0.11726   9.374  < 2e-16 ***
## size         0.50619    0.06644   7.619 2.56e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## size -0.054
```

Brown tree snakes

Gauss-Hermite quadrature

We can fit GHQ models with `lme4::glmer(..., nAGQ = pts)`

```
## fit model
snakes_ghq <- glmer(eggs ~ size + (1 | loc), data = df_eggs,
                    family = poisson, nAGQ = 20)
```

```
summary(snakes_ghq)
```

Note: this method only works with one random effect

Brown tree snakes

```
## Generalized linear mixed model fit by maximum likelihood (Adaptive
##   Gauss-Hermite Quadrature, nAGQ = 20) [glmerMod]
##  Family: poisson  ( log )
## Formula: eggs ~ size + (1 | loc)
##    Data: df_eggs
##
##      AIC      BIC   logLik deviance df.resid
##    397.7    408.1   -195.9    391.7      231
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1159 -0.8479 -0.2739  0.5929  4.0681
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  loc    (Intercept) 0.2761   0.5254
## Number of obs: 234, groups:  loc, 23
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.09919    0.11754   9.352  < 2e-16 ***
## size         0.50618    0.06681   7.576 3.56e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##      (Intr)
## size -0.054
```

Brown tree snakes

Here is a summary of the results from the 3 methods

```
##                  PQL    SE     Laplace   SE        GHQ    SE
## (Intercept)    1.125 0.117      1.099 0.117     1.099 0.118
## size           0.508 0.079      0.506 0.066     0.506 0.067
## location SD    0.508    NA      0.525    NA     0.525    NA
```

Brown tree snakes

What if we also want to include the random effect of year?

`glmmPQL` only allows for nested random effects

`glmer(..., nAGQ = pts)` only allows for one random effect

We can use the Laplace approximation via `glmer`

Brown tree snakes

Laplace

```
## fit model
snakes_lap_2 <- glmer(eggs ~ size + (1 | loc) + (1 | year),
                      data = df_eggs, family = poisson)
```

```
summary(snakes_lap_2)
```

Brown tree snakes

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: eggs ~ size + (1 | loc) + (1 | year)
##    Data: df_eggs
## 
##      AIC      BIC   logLik deviance df.resid 
##    928.8    942.6   -460.4    920.8      230 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.7498 -0.6251 -0.0568  0.5055  3.5431 
## 
## Random effects:
##  Groups Name        Variance Std.Dev.
##  loc    (Intercept) 0.2522   0.5022  
##  year   (Intercept) 0.1557   0.3945  
## Number of obs: 234, groups:  loc, 23; year, 14
## 
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)  1.03612    0.15518   6.677 2.44e-11 ***
## size         0.51380    0.07063   7.274 3.48e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##      (Intr)
## size -0.048
```

Diagnostics

Diagnostics for GLMMs are similar to those for GLMs, but we are limited in our choices

Goodness of fit

Recall our goodness of fit test based on the Pearson's χ^2

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(n-1)}$$

where O_i is the observed count and E_i is the expected count

Pearson's χ^2 statistic

For a binomial distribution

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i}$$

For a Poisson distribution

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \lambda_i)^2}{\lambda_i}$$

Goodness of fit

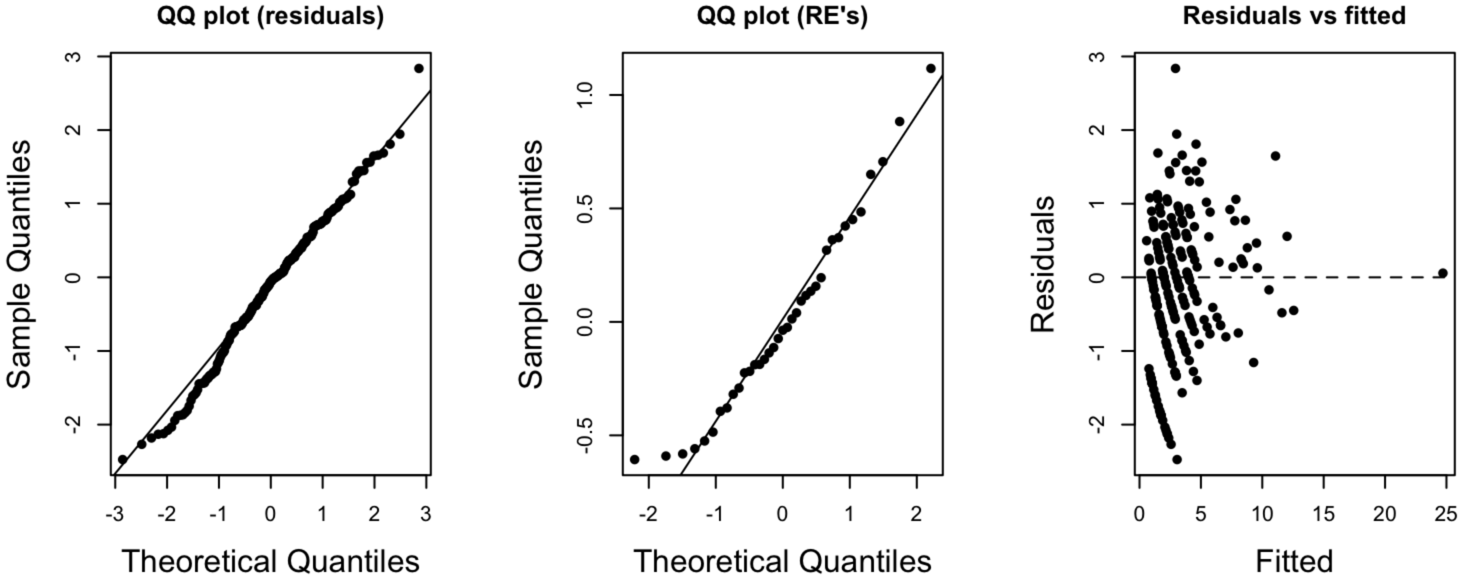
H_0: Our model is correctly specified

```r
## Pearson's X^2 statistic
X2 <- sum((eggs - fitted(snakes_lap_2))^2 / fitted(snakes_lap_2))
## likelihood ratio test
pchisq(X2, df = nn - length(coef(snakes_lap_2)),
       lower.tail = FALSE)
```

```
## [1] 0.9986993
```

The p-value is large so we cannot reject H_0

Model diagnostics



Leverage

For other models, we can calculate the leverages to evaluate potentially extreme values in predictor space

For GLMMs, however, the leverages depend on the estimated variance-covariance matrices of the random effects

Cook’s Distance

For other models, we can calculate Cook’s distances to identify potentially influential data points

For GLMMs, however, the Cook’s distances involve derivatives of the likelihood with respect to the random effects (this is an active area of research)

Inference for fixed effects

We can test the significance of the fixed effects via a χ^2 test by comparing models with and without the effect(s)

```
## fit reduced model
snakes_lap_null <- glmer(eggs ~ (1 | loc) + (1 | year),
                         data = df_eggs, family = poisson)
anova(snakes_lap_2, snakes_lap_null)
```

```
## Data: df_eggs
## Models:
## snakes_lap_null: eggs ~ (1 | loc) + (1 | year)
## snakes_lap_2: eggs ~ size + (1 | loc) + (1 | year)
##                 Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
## snakes_lap_null  3 979.77 990.14 -486.88   973.77
## snakes_lap_2     4 928.81 942.63 -460.40   920.81 52.961      1  3.402e-13
##
## snakes_lap_null
## snakes_lap_2    ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Inference for random effects

We can test the significance of the random effects via a χ^2 test by comparing models with and without the effect(s)

```r
## fit reduced model with only fixed effects
snakes_lap_null <- glm(eggs ~ size, data = df_eggs,
                       family = poisson(link = "log"))
## compare m0 and m1
anova(snakes_lap_2, snakes_lap_null)
```

```
## Data: df_eggs
## Models:
## snakes_lap_null: eggs ~ size
## snakes_lap_2: eggs ~ size + (1 | loc) + (1 | year)
##                 Df     AIC     BIC  logLik deviance  Chisq Chi Df
## snakes_lap_null  2 1173.92 1180.83 -584.96  1169.92
## snakes_lap_2     4  928.81  942.63 -460.40   920.81 249.11      2
##                 Pr(>Chisq)
## snakes_lap_null
## snakes_lap_2     < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Overdispersion

As with GLMs, we can check GLMMs for evidence of overdispersion, which we estimate as

$$\hat{c} = \frac{X^2}{n - k}$$

Let’s do so for our snake model applied to another data set

Overdispersion

```
## Pearson's X^2 statistic
X2 <- sum((eggs - fitted(snakes_lap))^2 / fitted(snakes_lap))
## number of parameters
k <- length(coef(snakes_lap)) + length(ranef(snakes_lap))
## overdispersion parameter
(c_hat <- X2 / (nn - k))
```

```
## [1] 2.767328
```

```
pchisq(deviance(snakes_lap), k, lower.tail = FALSE)
```

```
## [1] 5.191758e-216
```

Brown tree snakes

Negative binomial

We can fit neg binomial models using Laplace approximation with `lme4::glmer.nb()`

```
## fit model
snakes_lap_nb <- glmer.nb(eggs ~ size + (1 | loc) + (1 | year),
                          data = df_eggs)
```

```
summary(snakes_lap_nb)
```

Brown tree snakes

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: Negative Binomial(148955.1)  ( log )
## Formula: eggs ~ size + (1 | loc) + (1 | year)
##    Data: df_eggs
## 
##      AIC      BIC   logLik deviance df.resid 
##    930.8    948.1   -460.4    920.8      229 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -1.7498 -0.6251 -0.0568  0.5055  3.5431 
## 
## Random effects:
##  Groups Name        Variance Std.Dev.
##  loc    (Intercept) 0.2522   0.5022  
##  year   (Intercept) 0.1557   0.3946  
## Number of obs: 234, groups:  loc, 23; year, 14
## 
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept)  1.03607    0.15537   6.668 2.59e-11 ***
## size         0.51382    0.07122   7.215 5.41e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Correlation of Fixed Effects:
##      (Intr)
## size -0.048
```

Summary of GLMM methods

Method	Advantages	Disadvantages	R functions
Penalized quasi-likelihood	Flexible, widely implemented	inference may be inappropriate; potentially biased	`MASS::glmmPQL`

Laplace approximation	More accurate than PQL	Slower and less flexible than PQL	`lme4::glmer`
			`glmmsr::glmm`
			`glmmML::glmmML`

Gauss-Hermite quadrature	More accurate than Laplace	Slower than Laplace; limited random effects	`lme4::glmer`
			`glmmsr::glmm`
			`glmmML::glmmML`

Adapted from Bolker et al (2009)