R Package glmm: Likelihood-Based Inference for Generalized Linear Mixed Models

Christina Knudson, Ph.D.

University of St. Thomas

useR!2017
Reviewing the Linear Model

The usual linear model assumptions:

- responses normally distributed
- responses independent
- responses have equal variance
Extending the Linear Model

What if our response is not normal?

Use a generalized linear model and model

- the log odds of a dog following a command (binomial)
- the log mean number of dogs in a city block (Poisson)
What if the observations are correlated?

- repeated measures on one subject
- measurements on related/similar subjects

Use a linear mixed model and

- leave some parameters as “fixed effects”
- make others “random effects”
Why are they called “random” effects?

They are random variables, usually normal with mean 0.

Random effects are unobservable, but not parameters.

Variance component(s): variance(s) of the random effects.
Why are they called “random” effects?

They are random variables, usually normal with mean 0.

Random effects are unobservable, but not parameters.

Variance component(s): variance(s) of the random effects.

Parameters in LMM:

- fixed effects
- variance components
Extending the Linear Model

LMM assumptions
- responses: normally distributed, independent, and equal variance conditional on random effects
- random effects: normally distributed, independent, and mean zero (not necessarily same variance)
Extending the Linear Model

What if our observations are non-normal and correlated?

Use a “generalized linear mixed model” (GLMM)

- incorporate random effects to include correlation
- model log odds or log mean
Two salamander populations: Rough Butt (R) and White Side (W)
Do salamanders prefer mating with their own population?
Scientists reused salamanders in pairings:

Fred

Sally

Bob

Each salamander has a personalized tendency to mate.

We cannot measure this tendency.

We assume each salamander’s tendency is independent.
What affects the probability that a pair of salamanders mate?

- Type of cross (RR, RW, WR, WW)
- Female’s individualized tendency to mate
- Male’s individualized tendency to mate

The first is a fixed effect (average effect).

The next two are random effects (salamander-specific).
GLMM Example

Translation to statistical modeling:

- Response: whether or not the pair mated
- Fixed effects: $\beta_{RR}, \beta_{RW}, \beta_{WR}, \beta_{WW}$ (log odds of mating)
- Random effects: one per salamander, independent, normal
- Variance components: $\sigma_F^2$ and $\sigma_M^2$

How can we perform inference for GLMMS like this?
Likelihood: a function of the parameters given the observed data.

Likelihood-based inference includes:

- maximum likelihood (called “least squares” in LM)
- standard errors and covariances of parameter estimates
- confidence intervals
- hypothesis tests (Wald, LRT, etc)

Likelihood also used in AIC, BIC, etc
Inference for GLMMs

Why perform likelihood-based inference for GLMMs?

- MLE is asymptotically normal
- MLE’s cov matrix is a function of the likelihood’s Hessian

Likelihood: a function of the parameters given the observed data.

Likelihood-based inference is hard for GLMMs

- Likelihood cannot depend on random effects
- Likelihood is integral (often high dimension)
Inference for GLMMs

GLMM inference options:

- numerical integration enables likelihood-based inference for simple models (e.g. one random effect per observation)

- MCEM (performs maximum likelihood but no other likelihood-based inference)

- Penalized quasi-likelihood for approximate inference (lme4)

Released in 2015: R package glmm
(dissertation advisors Charles Geyer and Galin Jones)
R package glmm enables all likelihood-based inference:

- glmm approximates entire likelihood using Monte Carlo and importance sampling
- Monte Carlo MLEs converge to MLEs as $m \uparrow$
- Monte Carlo likelihood approximation converges to likelihood
- All likelihood-based inference converges
## Salamander Results

<table>
<thead>
<tr>
<th>Cross</th>
<th>RR</th>
<th>WW</th>
<th>RW</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of mating</td>
<td>0.736</td>
<td>0.730</td>
<td>0.584</td>
<td>0.126</td>
</tr>
</tbody>
</table>

(I’m skipping all the code. Download my slides to see it!)
How do point estimates compare?

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_{RR}$</th>
<th>$\hat{\beta}_{RW}$</th>
<th>$\hat{\beta}_{WR}$</th>
<th>$\hat{\beta}_{WW}$</th>
<th>$\hat{\nu}_F$</th>
<th>$\hat{\nu}_M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>glmm ($m = 10^6$)</td>
<td>1.03</td>
<td>.34</td>
<td>-1.94</td>
<td>1.00</td>
<td>1.36</td>
<td>1.23</td>
</tr>
<tr>
<td>MCEM</td>
<td>1.03</td>
<td>.32</td>
<td>-1.95</td>
<td>.99</td>
<td>1.4</td>
<td>1.25</td>
</tr>
<tr>
<td>lme4 (PQL)</td>
<td>1.01</td>
<td>.31</td>
<td>-1.89</td>
<td>.99</td>
<td>1.17</td>
<td>1.04</td>
</tr>
</tbody>
</table>
Inference with \texttt{glmm}

Model results outputted include:

- point estimates and standard errors
- likelihood, gradient, and Hessian

Additional \texttt{glmm} functions:

- Variance-covariance matrix (\texttt{vcov})
- Standard error (\texttt{se})
- Monte Carlo standard error (\texttt{mcse})
- Confidence intervals (\texttt{confint})
knudson.ust@gmail.com

cknudson.com

for slides and R package vignette
Comparing glmm and lme4

- *lme4* much faster (penalized-quasi likelihood v. Monte Carlo)
- *lme4* performs maximum likelihood for simple models (one random effect per observation)
- *glmm* performs/enables all likelihood-based inference
- *glmm* inference converges as \( m \uparrow \)
- *lme4* variance components are too small
- Impossible to know how close *lme4*’s PQL matches likelihood
- *glmm* currently limited to independent random effects
Salamander Example: Data Setup

> library(glmm)

> data(salamander)

> head(salamander)

<table>
<thead>
<tr>
<th>Mate</th>
<th>Cross</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>R/R</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>R/R</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>R/R</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>R/R</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>R/R</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>R/W</td>
<td>15</td>
</tr>
</tbody>
</table>
Salamander Example: Model Specification

> sal <- glmm(Mate ~ 0 + Cross,
  random = list( ~ 0 + Female, ~ 0 + Male ),
  varcomps.names = c( "F" , "M" ),
  data = salamander,  m = 10^6,
  family.glmm = binomial.glmm)

Notes:

- 0+Cross produces log odds for each group. Could use Cross if you want a reference group. (This is just like lm.)
- The random effects are centered at 0 almost always.
- Bigger m gives better estimates but takes more time.
Fixed Effects:

|                      | Estimate | Std. Error | z value |  Pr(>|z|) |
|----------------------|----------|------------|---------|-----------|
| CrossR/R             | 1.0253   | 0.4298     | 2.386   | 0.0170    * |
| CrossR/W             | 0.3375   | 0.3997     | 0.844   | 0.3984    |
| CrossW/R             | -1.9392  | 0.4694     | -4.131  | 3.61e-05  *** |
| CrossW/W             | 0.9961   | 0.4201     | 2.371   | 0.0177    * |

Familiar format (like `lm` summary)
Salamander Example: Model Summary

Variance Components for Random Effects
(P-values are one-tailed):

|    | Estimate | Std. Error | z     | Pr(>|z|)/2 |
|----|----------|------------|-------|------------|
| F  | 1.3647   | 0.6044     | 2.258 | 0.0120 *   |
| M  | 1.2331   | 0.6470     | 1.906 | 0.0283 *   |

---
Hypothesis Testing

We can translate the log odds back to probabilities:

\[ P(\text{mating}) = \frac{\exp(\hat{\beta}_{RW})}{1 + \exp(\hat{\beta}_{RW})} \]

<table>
<thead>
<tr>
<th>Cross</th>
<th>RR</th>
<th>WW</th>
<th>RW</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of mating</td>
<td>0.736</td>
<td>0.730</td>
<td>0.584</td>
<td>0.126</td>
</tr>
</tbody>
</table>
We can translate the log odds back to probabilities:

\[
P(\text{mating}) = \frac{\exp(\hat{\beta}_{RW})}{1 + \exp(\hat{\beta}_{RW})}
\]

<table>
<thead>
<tr>
<th>Cross</th>
<th>RR</th>
<th>WW</th>
<th>RW</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of mating</td>
<td>0.736</td>
<td>0.730</td>
<td>0.584</td>
<td>0.126</td>
</tr>
</tbody>
</table>

But which probabilities are significantly different?
Hypothesis Testing

Hypothesis tests determine which probabilities differ significantly.

\[ H_0 : \beta_{RR} = \beta_{WW} \]
\[ H_A : \beta_{RR} \neq \beta_{WW} \]

First, use `vcov` function for (co)variances needed to calculate

\[
\text{Var} \left( \hat{\beta}_{RR} - \hat{\beta}_{WW} \right) = \text{Var} \left( \hat{\beta}_{RR} \right) + \text{Var} \left( \hat{\beta}_{WW} \right) - 2 \text{Cov} \left( \hat{\beta}_{RR}, \hat{\beta}_{WW} \right).
\]

Then a Wald test statistic is

\[
\frac{\hat{\beta}_{RR} - \hat{\beta}_{WW} - 0}{\sqrt{\text{Var} \left( \hat{\beta}_{RR} - \hat{\beta}_{WW} \right)}} \sim \mathcal{N}(0, 1).
\]
Hypothesis Testing

Probability of mating does indeed depend on type of cross.

<table>
<thead>
<tr>
<th>Cross</th>
<th>RR</th>
<th>WW</th>
<th>RW</th>
<th>WR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of mating</td>
<td>0.736</td>
<td>0.730</td>
<td>0.584</td>
<td>0.126</td>
</tr>
</tbody>
</table>

An underline between two groups indicate the probabilities are not significantly different. For example, the odds of two rough butts mating are not significantly different from the odds of two white sides mating.
Monte Carlo Standard Error and \texttt{glmm}

\texttt{glmm} MCMLEs will vary from run to run (holding data constant).

This variability measured with Monte Carlo standard error:

\begin{verbatim}
> mcse(sal)

CrossR/R   CrossR/W   CrossW/R   CrossW/W
0.017468   0.032248   0.044544   0.024115

F        M
0.090671  0.055409
\end{verbatim}
Monte Carlo Standard Error and glmm

Compare two sources of variability:

- MCSE: variability from run to run, holding data constant
- SE: variability from data-set to data-set

If MCSE large compared to SE, increase $m$ to reduce MCSE. (Increasing $m$ will not decrease SE because data are fixed.)

```r
> se(sal)
CrossR/R   CrossR/W  CrossW/R   CrossW/W  ...
0.350252   0.366009  0.4222644  0.358033  ...
```


Related to PQL and lme4

*lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.1-6.


R package glmm

1. Based on data, selects importance sampling distribution $\tilde{f}(u)$
2. Generates $m$ random effects from $\tilde{f}(u)$
3. Calculates and maximizes MCLA using trust
4. Returns
   - Monte Carlo MLEs
   - MCLA value, gradient and Hessian at MCMLEs
   - Lots of other info (trust output, etc)

Families currently allowed: Binomial and Poisson

Random effect structure currently allowed: independent normals