Generalized linear mixed models

Ben Bolker

McMaster University, Mathematics & Statistics and Biology

30 June 2015

Acknowledgments

- `lme4`: Doug Bates, Martin Mächler, Steve Walker
- Data: Josh Banta, Adrian Stier, Sea McKeon, David Julian, Jada-Simone White

- NSERC (Discovery)
- SHARCnet

Outline

1. Examples and definitions
2. Estimation
    - Overview
    - Methods
3. Inference
4. Challenges & open questions

Outline

(Generalized) linear mixed models

(G)LMMs: a statistical modeling framework incorporating:

- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses (e.g. binomial, Poisson, and extensions)
- (some) nonlinearity (e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data

(Generalized) linear mixed models

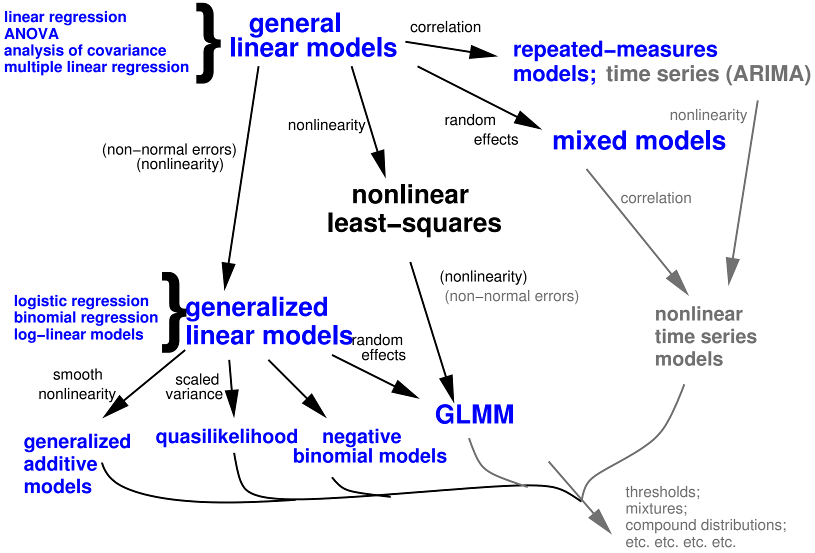
(G)LMMs: a statistical modeling framework incorporating:

- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses (e.g. binomial, Poisson, and extensions)
- (some) nonlinearity (e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data

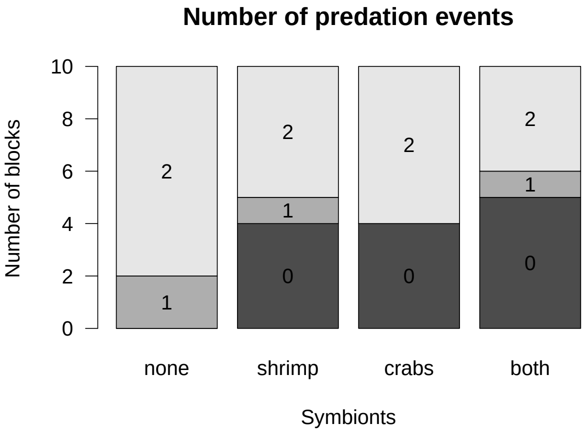
(Generalized) linear mixed models

(G)LMMs: a statistical modeling framework incorporating:

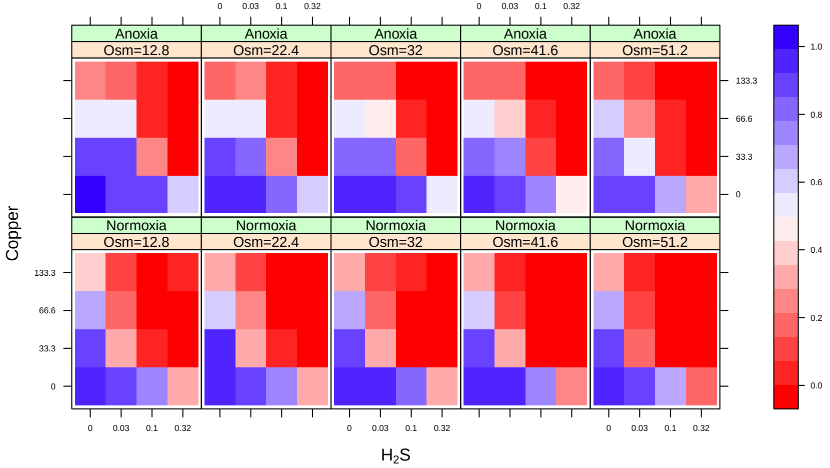
- combinations of categorical and continuous predictors, and interactions
- (some) non-Normal responses (e.g. binomial, Poisson, and extensions)
- (some) nonlinearity (e.g. logistic, exponential, hyperbolic)
- non-independent (grouped) data


linear regression
ANOVA
analysis of covariance
multiple linear regression
general linear models
correlation
repeated–measures models; time series (ARIMA)
random effects
mixed models
nonlinearity
nonlinearity
(non–normal errors)
(nonlinearity)
nonlinear least–squares
correlation
(nonlinearity)
(non–normal errors)
logistic regression
binomial regression
log–linear models
generalized linear models
random effects
nonlinear time series models
smooth nonlinearity
scaled variance
GLMM
generalized additive models
quasilikelihood
negative binomial models
thresholds;
mixtures;
compound distributions;
etc. etc. etc. etc.

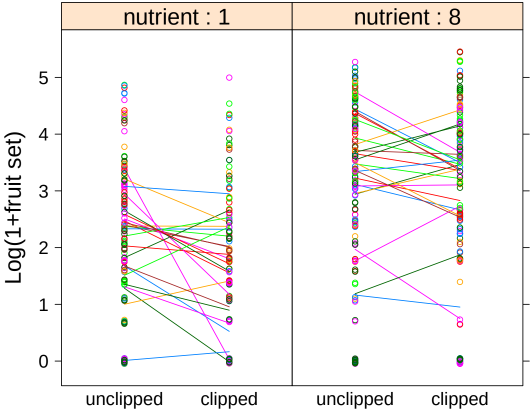
Coral protection from seastars (*Culcita*) by symbionts (McKeon et al., 2012)



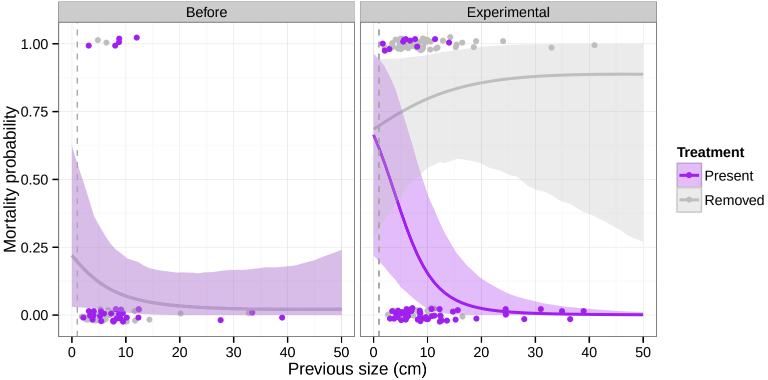
Environmental stress: *Glycera* cell survival (D. Julian unpubl.)



Arabidopsis response to fertilization & herbivory (Banta et al., 2010)



Coral demography (J.-S. White unpubl.)



Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\substack{\text{conditional}\\ \text{distribution}}} (\underbrace{g^{-1}(\eta_i)}_{\substack{\text{inverse}\\ \text{link}\\ \text{function}}}, \underbrace{\phi}_{\substack{\text{scale}\\ \text{parameter}}})$$

Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\substack{\text{conditional}\\\text{distribution}}} (\underbrace{g^{-1}(\eta_i)}_{\substack{\text{inverse}\\\text{link}\\\text{function}}}, \underbrace{\phi}_{\substack{\text{scale}\\\text{parameter}}})$$

$$\underbrace{\boldsymbol{\eta}}_{\substack{\text{linear}\\\text{predictor}}} = \underbrace{\boldsymbol{X\beta}}_{\substack{\text{fixed}\\\text{effects}}} + \underbrace{\boldsymbol{Zb}}_{\substack{\text{random}\\\text{effects}}}$$

Technical definition

$$\underbrace{Y_i}_{\text{response}} \sim \overbrace{\text{Distr}}^{\substack{\text{conditional}\\\text{distribution}}} (\underbrace{g^{-1}(\eta_i)}_{\substack{\text{inverse}\\\text{link}\\\text{function}}}, \underbrace{\phi}_{\substack{\text{scale}\\\text{parameter}}})$$

$$\underbrace{\boldsymbol{\eta}}_{\substack{\text{linear}\\\text{predictor}}} = \underbrace{\boldsymbol{X\beta}}_{\substack{\text{fixed}\\\text{effects}}} + \underbrace{\boldsymbol{Zb}}_{\substack{\text{random}\\\text{effects}}}$$

$$\underbrace{\boldsymbol{b}}_{\substack{\text{conditional}\\\text{modes}}} \sim \text{MVN}(\mathbf{0}, \underbrace{\Sigma(\boldsymbol{\theta})}_{\substack{\text{variance-}\\\text{covariance}\\\text{matrix}}})$$

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and fixed effects (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for ...

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for ...

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for ...

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

What are random effects?

A method for . . .

- accounting for among-individual, within-block correlation
- compromising between
  complete pooling (no among-block variance)
  and **fixed effects** (large among-block variance)
- handling levels selected at random from a larger population
- sharing information among levels (*shrinkage estimation*)
- estimating variability among levels
- allowing predictions for unmeasured levels

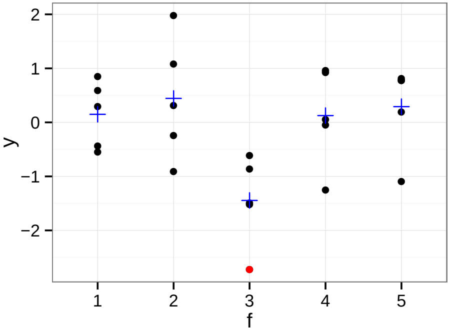
Random-effect myths

- levels of random effects must always be sampled at random
- a complete sample cannot be treated as a random effect
- random effects are always a **nuisance variable**
- nothing can be said about the predictions of a random effect
- you should always use a random effect no matter how few levels you have

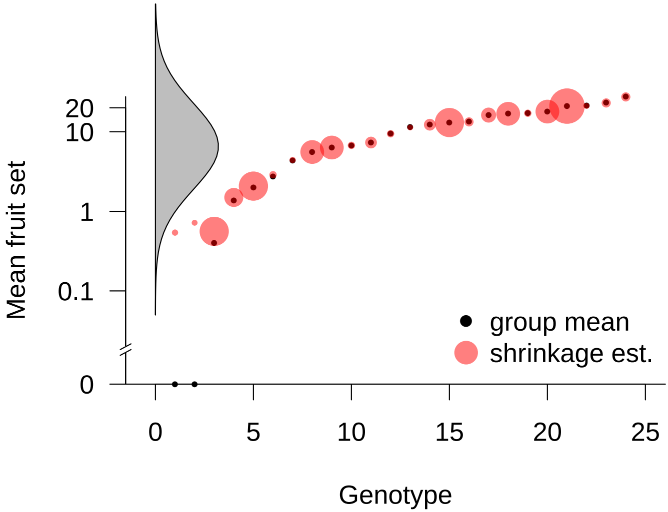
Outline

Maximum likelihood estimation

- Best fit is a compromise between two components (consistency of data with fixed effects and conditional modes; consistency of random effect with RE distribution)
- Goodness-of-fit *integrates* over conditional modes



Shrinkage: *Arabidopsis* conditional modes



Estimation methods

deterministic : various approximate integrals (Breslow, 2004)

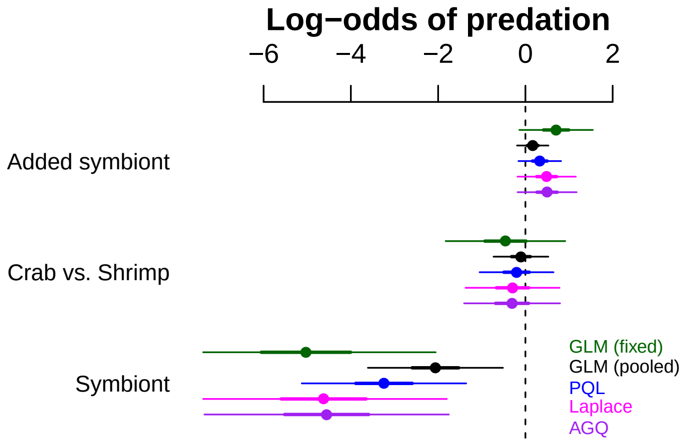
- Penalized quasi-likelihood, Laplace, Gauss-Hermite quadrature, . . . (?); best methods needed for large variance, small clusters
- flexibility and speed vs. accuracy

. . .

stochastic (Monte Carlo): frequentist and Bayesian (Booth and Hobert, 1999; Ponciano et al., 2009; Sung, 2007)

- usually slower but flexible and accurate

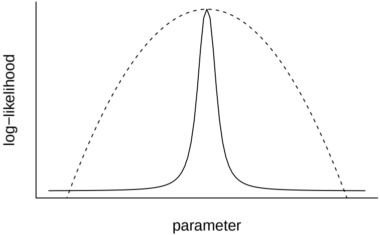
Estimation: *Culcita* (McKeon et al., 2012)



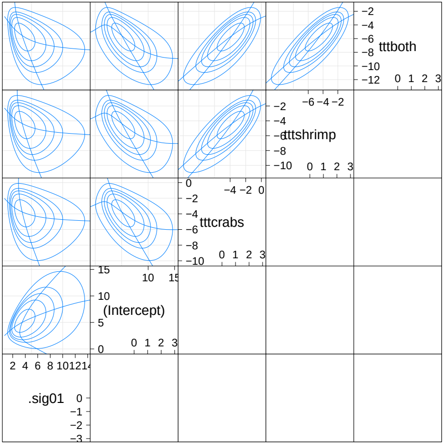
Outline

Wald tests

- typical results of `summary`
- exact for ANOVA, regression: approximation for GLM(M)s
- fast
- approximation is sometimes awful (Hauck-Donner effect)



2D profiles for *Culcita* data



Scatter Plot Matrix

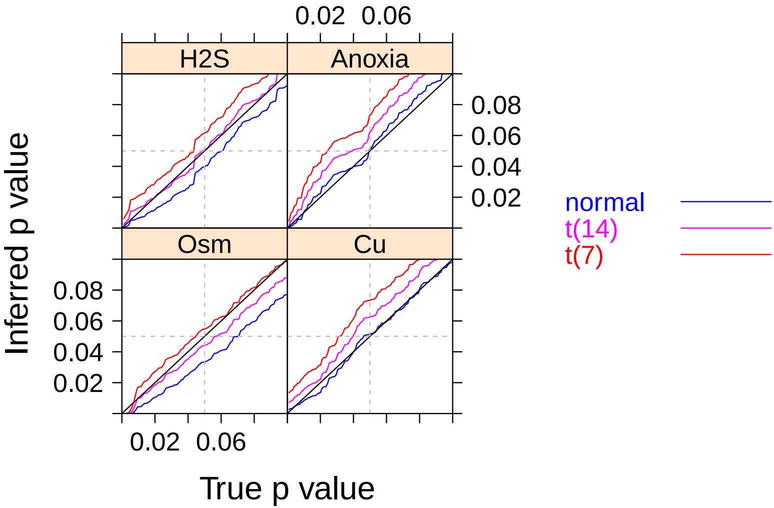
Likelihood ratio tests

- better than Wald, but still have to two problems:
  - "denominator degrees of freedom" (when estimating scale)
  - for GLMMs, distributions are approximate anyway (Bartlett corrections)
  - Kenward-Roger correction? (Stroup, 2014)
- Profile confidence intervals: expensive/fragile

Parametric bootstrapping

- fit null model to data
- simulate "data" from null model
- fit null and working model, compute likelihood difference
- repeat to estimate null distribution
- should be OK but ??? not well tested
  (assumes estimated parameters are "sufficiently" good)

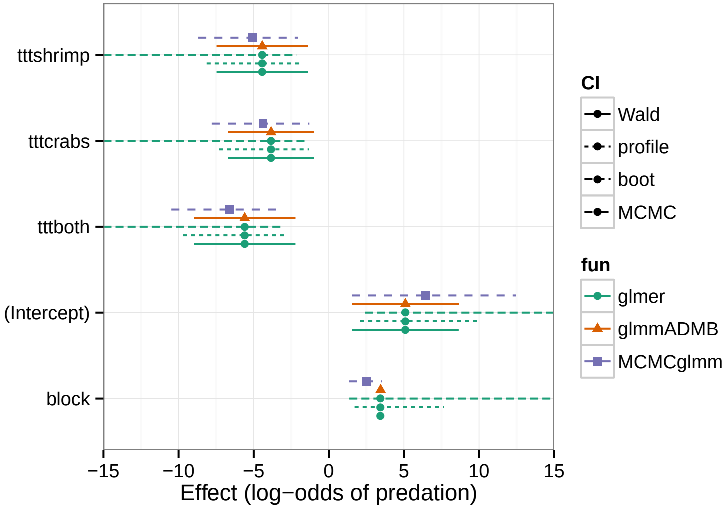
Parametric bootstrap results (*Glycera*)



Bayesian inference

- If we have a good sample from the posterior distribution (Markov chains have converged etc. etc.) we get most of the inferences we want for free by summarizing the marginal posteriors
- *post hoc* Bayesian can work, but mode at zero causes problems

Culcita confidence intervals



Outline

On beyond R

- Julia: MixedModels package
- SAS: PROC MIXED, NLMIXED
- AS-REML
- Stata (GLLAMM, xtmelogit)
- AD Model Builder; Template Model Builder
- HLM, MLWiN

Challenges

- Small clusters: need AGQ/MCMC
- Small numbers of clusters: need finite-size corrections (KR/PB/MCMC)
- Small data sets: issues with **singular** fits
  Barr et al. (2013) vs. Bates et al. (2015)
- Big data: speed!
- Model diagnosis
- Confidence intervals accounting for uncertainty in variances

See also: `http://rpubs.com/bbolker/glmmchapter`, `https://groups.nceas.ucsb.edu/non-linear-modeling/projects`

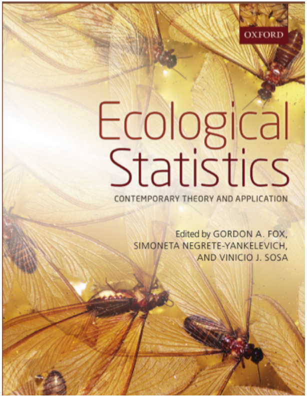
Spatial and temporal correlations

- Sometimes blocking takes care of non-independence ...
- but sometimes there is temporal or spatial correlation **within** blocks
- . . . also phylogenetic . . . (Ives and Zhu, 2006)
- “G-side” vs. “R-side” effects
- tricky to implement for GLMMs,
  but new possibilities on the horizon (Rousset and Ferdy, 2014; Rue et al., 2009);
  https://github.com/stevencarlislewalker/lme4ord

Next steps

- Complex random effects:
  regularization, model selection, penalized methods (lasso/fence)
- Flexible correlation and variance structures
- Flexible/nonparametric random effects distributions
- hybrid & improved MCMC methods
- **Reliable** assessment of out-of-sample performance

- URL here
- http://www.math.mcmaster.ca/bolker/R/misc/foxchapter
- Bolker (2015)



References

Banta, J.A., Stevens, M.H.H., and Pigliucci, M., 2010. *Oikos*, 119(2):359–369. ISSN 1600-0706. doi:10.1111/j.1600-0706.2009.17726.x.

Barr, D.J., Levy, R., et al., 2013. *Journal of Memory and Language*, 68(3):255–278. ISSN 0749-596X. doi:10.1016/j.jml.2012.11.001.

Bates, D., Kliegl, R., et al., 2015. *arXiv:1506.04967 [stat]*. ArXiv: 1506.04967.

Bolker, B.M., 2015. In G.A. Fox, S. Negrete-Yankelevich, and V.J. Sosa, editors, *Ecological Statistics: Contemporary theory and application*. Oxford University Press. ISBN 978-0-19-967255-4.

Booth, J.G. and Hobert, J.P., 1999. *Journal of the Royal Statistical Society. Series B*, 61(1):265–285. doi:10.1111/1467-9868.00176.

Breslow, N.E., 2004. In D.Y. Lin and P.J. Heagerty, editors, *Proceedings of the second Seattle symposium in biostatistics: Analysis of correlated data*, pages 1–22. Springer. ISBN 0387208623.

Ives, A.R. and Zhu, J., 2006. *Ecological Applications*, 16(1):20–32.

McKeon, C.S., Stier, A., et al., 2012. *Oecologia*, 169(4):1095–1103. ISSN 0029-8549. doi:10.1007/s00442-012-2275-2.

Ponciano, J.M., Taper, M.L., et al., 2009. *Ecology*, 90(2):356–362. ISSN 0012-9658.

Rousset, F. and Ferdy, J.B., 2014. *Ecography*, page no–no. ISSN 1600-0587. doi:10.1111/ecog.00566.

Rue, H., Martino, S., and Chopin, N., 2009. *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Stroup, W.W., 2014. *Agronomy Journal*, 106:1–17. doi:10.2134/agronj2013.0342.

Sung, Y.J., 2007. *The Annals of Statistics*, 35(3):990–1011. ISSN 0090-5364. doi:10.1214/009053606000001389.