Workshop 11.2a: Generalized Linear Mixed Effects Models (GLMM)

Murray Logan

07 Feb 2017

Section 1

Generalized Linear Mixed Effects Models

Parameter Estimation

lm ?? LME

(integrate likelihood across all unobserved levels random effects)

Parameter Estimation

lm □□ LME

(integrate likelihood across all unobserved levels random effects)

glm □-□□□□□ GLMM

Not so easy - need to approximate

Parameter Estimation

- Penalized quasi-likelihood
- Laplace approximation
- Gauss-Hermite quadrature

Penalized quasi-likelihood (PQL)

ITERATIVE (RE)WEIGHTING

- LMM to estimate vcov structure
- fixed effects estimated by fitting GLM (incorp vcov)
- refit LMM to re-estimate vcov
- cycle

Penalized quasi-likelihood (PQL)

ADVANTAGES

- relatively simple
- leverage variance-covariance structures for heterogeneity and dependency structures

DISADVANTAGES

- biased when expected values less ⍰5
- approximates likelihood (no AIC or LTR)

Laplace approximation

Second-order Taylor series expansion - to approximate likelihood at unobserved levels of random effects

Laplace approximation

Second-order Taylor series expansion - to approximate likelihood at unobserved levels of random effects

ADVANTAGES

- more accurate

Laplace approximation

Second-order Taylor series expansion - to approximate likelihood at unobserved levels of random effects

ADVANTAGES

- more accurate

DISADVANTAGES

- slower
- no way to incorporate vcov

Gauss-Hermite quadrature (GHQ)

- approximates value of integrals at specific points (quadratures)
- points (and weights) selected by optimizer

Gauss-Hermite quadrature (GHQ)

- approximates value of integrals at specific points (quadratures)
- points (and weights) selected by optimizer

ADVANTAGES

- even more accurate

Gauss-Hermite quadrature (GHQ)

- approximates value of integrals at specific points (quadratures)
- points (and weights) selected by optimizer

ADVANTAGES

- even more accurate

DISADVANTAGES

- even slower
- no way to incorporate vcov

Markov Chain Monte Carlo (MCMC)

- recreate likelihood by sampling proportionally to likelihood

Markov Chain Monte Carlo (MCMC)

- recreate likelihood by sampling proportionally to likelihood

ADVANTAGES

- very accurate (not an approximation)
- very robust

Markov Chain Monte Carlo (MCMC)

- recreate likelihood by sampling proportionally to likelihood

ADVANTAGES

- very accurate (not an approximation)
- very robust

DISADVANTAGES

- very slow

Inference (hypothesis) testing

GLMM

Depends on:

- Estimation engine (PQL, Laplace, GHQ)
- Overdispersed
- Fixed or random factors

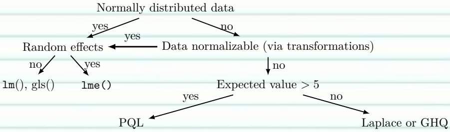
Inference (hypothesis) testing

Approximation	Characteristics	Associated inference	R Function
Penalized Quasi-likelihood (PQL)	Fast and simple, accommodates heterogeneity and dependency structures, biased for small samples	Wald tests only	**glmmPQL** (MASS)
Laplace	More accurate (less biased), slower, does not accommodate heterogeneity and dependency structures	LRT	**glmer** (lme4), **glmmadmb** (glmmADMB)
Gauss-Hermite quadrature	Evan more accurate (less biased), slower, does not accommodate heterogeneity and dependency structures, cant handle more than 1 random effect	LRT	**glmer** (lme4)?? - does not seem to work
Markov Chain Monte Carlo (MCMC)	Bayesian, very flexible and accurate, yet very slow and more complex	Bayesian credibility intervals, Bayes factors	Numerous (see Tutorial 9.2b)

Inference (hypothesis) testing

Feature	**glmmQPL** (MASS)	**glmer** (lme4)	**glmmadmb** (glmmADMB)	MCMC
Varoamce amd covariance structures	Yes	-	not yet	Yes
Overdispersed (Quasi) families	Yes	limited	some	-
Mixture families	limited	limited	limited	Yes
Zero-inflation	-	-	Yes	Yes
Residual degrees of freedom	Between-within	-?	-	NA
Parameter tests	Wald t	Wald Z	Wald Z	UI
Marginal tests (fixed effects)	Wald F, χ^2	Wald F, χ^2	Wald F, χ^2	UI
Marginal tests (random effects)	Wald F, χ^2	LRT	LRT	UI
Information criterion	-	AIC	AIC	AIC, WAIC

Inference (hypothesis) testing



Overdispersed	Model	Inference
No	`glmmPQL()`	Wald Z or χ^2
Yes	`glmmPQL(.., family='quasi..')`	Wald t or F
Clumpiness	`glmmPQL(.., family='negative.binomial')`	Wald t or F
Zero-inflation	`glmmadmb(.., zeroInflated=TRUE)`	Wald t or F

Overdispersed	Model	Inference
Random effects		
Yes or no	`glmer()` or `glmmadmb()`	LRT (ML)
Fixed effects		
No	`glmer()` or `glmmadmb()`	Wald Z or χ^2
Yes	`glmer(..(1|Obs))`	Wald t or F
Clumpiness	`glmer(.., family='negative.binomial')`	Wald t or F
	`glmmamd(.., family='nbinom')`	Wald t or F
Zero-inflation	`glmmadmb(.., zeroInflated=TRUE)`	Wald t or F

Additional assumptions

- dispersion
- (multi)collinearity
- design balance and Type III (marginal) SS
- heteroscadacity
- spatial/temporal autocorrelation

Section 2

Worked Examples

Worked Examples

$$\mathtt{log}(\mathtt{y_{ij}}) = \gamma_{\mathtt{Site_i}} + \beta_0 + \beta_1 \mathtt{Treat_i} + \varepsilon_{\mathtt{ij}} \qquad \varepsilon \sim \mathtt{Pois}(\lambda)$$

$$\mathtt{where} \sum \gamma = 0$$