Estimation Methods for Generalized Linear Mixed Models with Binary Outcomes from Small Clusters

10th International Multilevel Conference

Haeike Josephy - Tom Loeys

Department of Data-Analysis, Ghent University - Thursday 9th April, 2015

Overview

Goal

Methods

Simulations

Results

Conclusion

Goal of this presentation

- Consider multilevel data with binary outcome measures
  → e.g. Generalised Linear Mixed Models (GLMMs)
- Consider cluster size two
  → e.g. crossover studies, dyadic data, ...
- Compare the performance of different appropriate methods
  → Assess several available functions in R (R Core Team, 2013)

Why?

Settings with binary outcomes in small clusters have proven difficult for the available methodologies (Xu et al., 2014)

Goal of this presentation

- Consider multilevel data with binary outcome measures
  → e.g. Generalised Linear Mixed Models (GLMMs)
- Consider cluster size two
  → e.g. crossover studies, dyadic data, ...
- Compare the performance of different appropriate methods
  → Assess several available functions in R (R Core Team, 2013)

Why?

Settings with binary outcomes in small clusters have proven difficult for the available methodologies (Xu et al., 2014)

Goal of this presentation

- Consider multilevel data with binary outcome measures
  → e.g. Generalised Linear Mixed Models (GLMMs)
- Consider cluster size two
  → e.g. crossover studies, dyadic data, ...
- Compare the performance of different appropriate methods
  → Assess several available functions in R (R Core Team, 2013)

Why?

Settings with binary outcomes in small clusters have proven difficult for the available methodologies (Xu et al., 2014)

Goal of this presentation

- Consider multilevel data with binary outcome measures
  → e.g. Generalised Linear Mixed Models (GLMMs)
- Consider cluster size two
  → e.g. crossover studies, dyadic data, ...
- Compare the performance of different appropriate methods
  → Assess several available functions in R (R Core Team, 2013)

Why?

Settings with binary outcomes in small clusters have proven difficult for the available methodologies (Xu et al., 2014)

General difficulties

GLMMs are most widespread for handling binary multilevel data, BUT:

- Statistical inference of GLMMs is hampered due to its random effects (RE's):
    - Likelihood function involves integrating out these effects from the joint density of responses and RE's
    - This is (except for a few cases) analytically intractable
- To tackle this intractability, numerous estimation methods have been proposed:
    - Likelihood-based approximation methods
    - Bayesian estimation procedures
    - Least Squares (LS) procedures in the Structural Equation Modelling (SEM) framework

General difficulties

GLMMs are most widespread for handling binary multilevel data, BUT:

- Statistical inference of GLMMs is hampered due to its random effects (RE's):
  - Likelihood function involves integrating out these effects from the joint density of responses and RE's
  - This is (except for a few cases) analytically intractable
- To tackle this intractability, numerous estimation methods have been proposed:
  - Likelihood-based approximation methods
  - Bayesian estimation procedures
  - Least Squares (LS) procedures in the Structural Equation Modelling (SEM) framework

Solutions (1/3)

Likelihood-based approximation methods:

1. Laplace approximation
   - Approximates the intractable integrand by a quadratic Taylor expansion

     $\rightarrow$ Closed-form expression of the maximizable likelihood
   - in R: `glmer` (package `lme4`)
2. Penalised Quasi-Likelihood method (PQL)

   (Breslow and Clayton, 1993; Schall, 1991; Stiratelli et al., 1984)
   - Also an approximation of the integrand
   - Considered an approximation of the GLMM by a LMM

     $\rightarrow$ estimation simplifies
   - in R: `glmmPQL` (package `MASS`)

Solutions (1/3)

Likelihood-based approximation methods:

1. Laplace approximation
   - Approximates the intractable integrand by a quadratic Taylor expansion
     → Closed-form expression of the maximizable likelihood
   - in R: `glmer` (package `lme4`)
2. Penalised Quasi-Likelihood method (PQL)
   (Breslow and Clayton, 1993; Schall, 1991; Stiratelli et al., 1984)
   - Also an approximation of the integrand
   - Considered an approximation of the GLMM by a LMM
     → estimation simplifies
   - in R: `glmmPQL` (package `MASS`)

Solutions (1/3)

Likelihood-based approximation methods:

3. Adaptive Gaussian Quadrature (AGQ)(Pinheiro and Bates, 1995)
   - Approximates the integral by replacing it with a finite sum:
     - regular Gauss-Hermite (GH) quadrature (e.g. (Naylor and Smith, 1982)) uses fixed set of nodes
     - AGQ uses a different set of nodes for each cluster.
       $\rightarrow$ more efficient than GH quadrature
   - in R: `glmer` (package lme4, option 'nAGQ>1')

Solutions (2/3)

Bayesian estimation procedures:

4. Markov Chain Monte Carlo (MCMC) methods
   - Simulate the likelihood, rather than computing it
     → Calculate sample average of independently simulated realisations of the integrand
   - in R: `MCMCglmm` (package `MCMCglmm`)

5. Hybrid approach
   - Uses an Integrated Nested Laplace Approximation (INLA) of the posterior distributions

     → No need to simulate the likelihood

     → Steep decline in computational burden
   - in R: `inla` (package `R-inla`)

Solutions (2/3)

Bayesian estimation procedures:

4. Markov Chain Monte Carlo (MCMC) methods
   - Simulate the likelihood, rather than computing it
     → Calculate sample average of independently simulated realisations of the integrand
   - in R: `MCMCglmm` (package `MCMCglmm`)

5. Hybrid approach
   - Uses an Integrated Nested Laplace Approximation (INLA) of the posterior distributions

     → No need to simulate the likelihood

     → Steep decline in computational burden
   - in R: `inla` (package `R-inla`)

Solutions (3/3)

LS estimation in SEM:

- Different estimation techniques available:
  - OLS, DWLS, GLS
  - Diagonally Weighted Least Squares (DWLS)
    $\rightarrow$ more robust and accurate than OLS, GLS only for $n > 10000$.
    $\rightarrow$ only for probit link
- SEM theoretical background:
  - Clustered binary outcome Y_{ij} represents crude approximation of underlying continuous variable $\widetilde{Y}_{ij}$.
  - $\widetilde{Y}_{ij}$ is not directly observed (*latent*), where:

$$\widetilde{Y}_{ij} = \beta_0 + \beta_1 x_{ij} + b_j + \epsilon_{ij} \tag{1}$$

, with ϵ_{ij} the residual variance $\sim N(0, \sigma^2)$ and b_j a random intercept $\sim N(0, \tau)$.

  - $Y_{ij} = 1 \iff \widetilde{Y}_{ij} > c$, with c a threshold value

Solutions (3/3)

LS estimation in SEM:

- Two parameterisations:
  6. In traditional literature, σ^2 is fixed at one ($\epsilon_{ij} \sim N(0,1)$)
     $\Rightarrow$ Theta approach (Muthén and Muthén, 2010).
  7. In SEM literature $\tau + \sigma^2$ is fixed at one ($b_j + \epsilon_{ij} \sim N(0,1)$)
     $\Rightarrow$ Delta approach (Muthén and Muthén, 2010).
- They provide different estimates, convertible by a scaling factor Δ (Muth et al., 2002) (here, $\Delta = 1/\sqrt{(\tau+1)}$).
- in R: `sem` (package `lavaan`, with option '`parameterization=theta/delta`')

Simulations

Simulate binary outcome data of cluster size two, generated with a probit link. We look at different settings for:

- Sample size: $n = 25, 50, 100, 500$
- Intracluster correlation: $ICC = 0.1, 0.3, 0.5$
- Measure for the exposure X:
  - Binary/Gaussian (*scale*)
  - Between-/Within- cluster (*bw*)
- Event rate: $P(Y = 1) = 0.5$ (0.1 in progress)

$\Rightarrow$ Compare all seven methods in terms of:

Bias - SE - MSE - Coverage - Convergence

Simulations

Simulate binary outcome data of cluster size two, generated with a probit link. We look at different settings for:

- Sample size: $n = 25, 50, 100, 500$
- Intracluster correlation: $ICC = 0.1, 0.3, 0.5$
- Measure for the exposure X:
  - Binary/Gaussian (*scale*)
  - Between-/Within- cluster (*bw*)
- Event rate: $P(Y = 1) = 0.5$ (0.1 in progress)

$\Rightarrow$ Compare all seven methods in terms of:

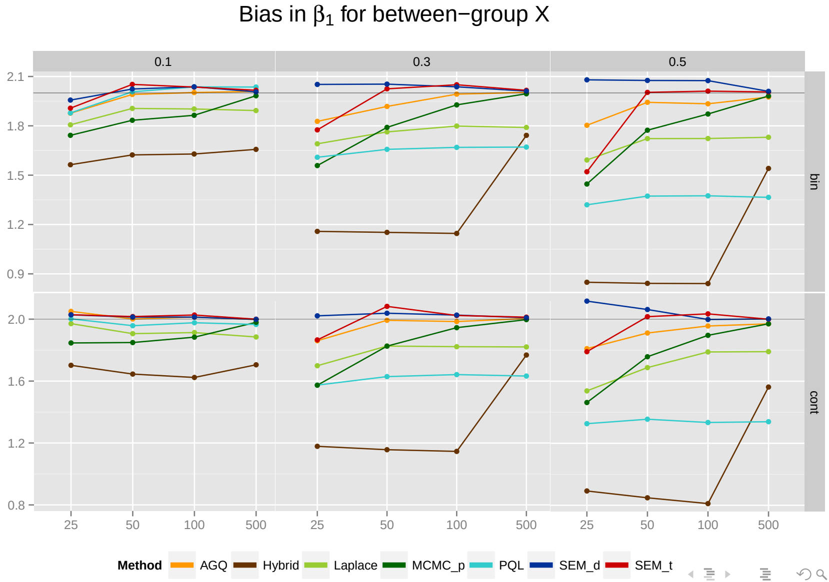
Bias - SE - MSE - Coverage - Convergence

Results for β_1

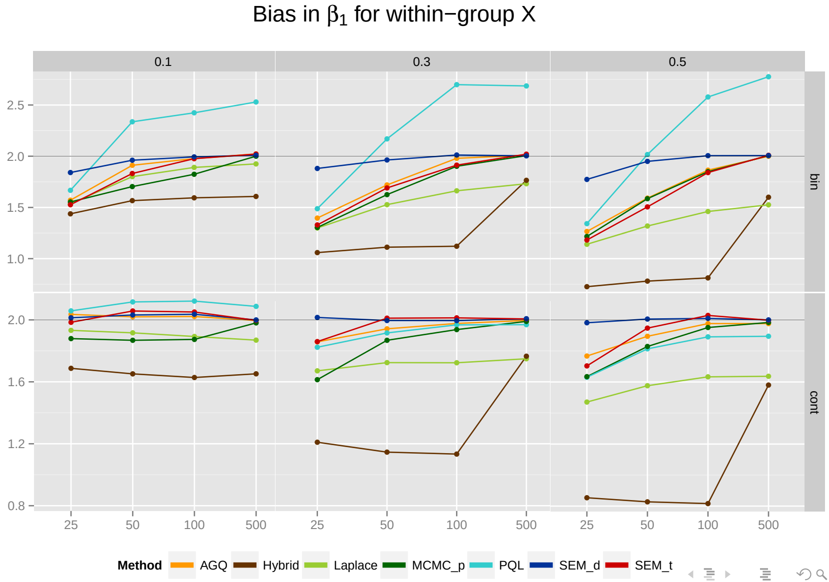
Method	Bias	Stand. error	MSE	Coverage	Convergence
Laplace	bw*scale*n	bw*scale*icc	bw*scale*icc	icc + scale	n
AGQ	bw*scale*n	bw*icc	bw*icc	n + icc	n + bw
PQL	bw*scale*n	bw*scale*n	bw*scale*n	n + icc + bw	bw*scale*n
MCMC	bw*scale*n	icc*n bw*scale	icc*n bw*scale	n	icc*n + bw*icc bw*n + bw*scale
Hybrid	bw*scale*n	scale*n + bw*n bw*scale	icc*n	icc*n	/
SEM-δ	scale*n + bw*n bw*scale	n + icc scale + bw	n + icc	scale*icc + bw*icc bw*scale	scale*icc*n
SEM-θ	scale*n + bw*n bw*scale	scale*n + bw*n bw*bin	icc*n + scale*n bw*n	bw*scale*n	bw*scale*n

*significant terms at the 0.14% significance level

Results for β_1



Results for β_1



Conclusion

- Testing for factors (n, icc, bw & scale)
    - Each factor is relevant!
    - BUT some methods show more variance than others.

      $\rightarrow$ Nonsensical to compare methods in terms of significant factors

      $\rightarrow$ Required: elegant way to compare all approaches...
- Graphical comparison
    - For bias of β_1: does seem to favour SEM-δ
    - SEM-δ also performs well for SE, MSE and convergence
- Additional method: Pairwise Maximum Likelihood
    - One-step estimation $\rightarrow$ probably more efficient than LS
    - Recently implemented in `lavaan`
- Additional factor: outcome prevalence

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized lineat mixed models. *Journal of the American Statistical Association*, 88(421):9–25.

Muth, B., Muthén, B., and Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus web notes*, 4(5):0–22.

Muthén, L. K. and Muthén, B. O. (2010). *Mplus User's Guide*. Muthén & Muthén, Los Angeles, CA, sixth edit edition.

Naylor, J. and Smith, A. (1982). Applications of a method for the efficients computation of posterior distributions. *Applied Statistics*, 31(3):214–225.

Pinheiro, J. and Bates, D. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4:12–35.

R Core Team (2013). R: A Language and Environment for Statistical Computing.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–727.

Stiratelli, R., Laird, N., and Ware, J. (1984). Random-effects models for serial observations with binary response. *Biometrics*, 40(4):961–971.

Xu, Y., Lee, C. F., and Cheung, Y. B. (2014). Analyzing Binary Outcome Data with Small Clusters: A Simulation Study. *Communications in Statistics - Simulation and Computation*, 43(March 2015):1771–1782.