**Estimation Methods for Non-continuous Multilevel Regression**

| Estimation Method | Software | Algorithms | Comments |
|---|---|---|---|
| Penalized quasi-likelihood (PQL) | Used by HLM for binomial and poisson models, but Raudenbush and Bryk (2002) recommend combination with Laplace approximation. One of the available options in SAS. SPSS offers only the basic PQL approach in the GENLINMIXED procedure (beginning with Version 19). | Fixed and random effects are included in Taylor expansion. May be used as starting values for Gaussian quadrature approach. | Can be seriously biased if high level random effects are large and outcome probability is very low or very high. Fixed effects, random effects, and standard errors generally too low. Better estimates with larger level-1 sample sizes. Performs much better than older marginal quasi-likelihood (MQL) approach, however. |
| Restricted or residualized penalized quasi-likelihood (RPQL) | Only available currently in `SAS PROC GLIMMIX` | A modified process similar to the REML difference from full ML. | Performance is much better than standard PQL. Work well with small number of clusters (and outperformers adaptive quadrature and Laplace), especially using Kenward-Roger corrections and with large number of cases per group (McNeish, 2016, 2019; McNeish & Stapleton, 2016). |
| High-order Laplace approximation (Raudenbush, Yang, & Yosef, 2000) | HLM software's preferred method "Laplace6" for 2 and 3-level Bernoulli and 2-level Poisson models. Available in SAS 9.2 `PROC GLIMMIX` when `METHOD=LAPLACE` is included on the proc line. R `lmer` function in the `lme4` package has a variant of the laplace estimation used in HLM that is equivalent to the adaptive quadrature with only one integration point (estimated by default with `family = binary`). | HLM uses PQL Taylor expansion approach plus a second iterative process (micro and macro iterations). The HLM package approach uses a sixth-order Laplace approximation to maximize the likelihood (Laplace6; Raudenbush, Yang, & Yosef, 2000) to avoid biases associated with PQL. Fisher scoring is used for maximization. Newton-Raphson iteration and Fisher scoring is equivalent to the Gaussian approach with one quadrature point (Wolfinger, 1993). The Laplace estimation in `lme4` does not use a higher-order approximation and is not equivalent. | Provides highly accurate estimates under a variety of conditions. Seems to perform well in many situations. Can be used for likelihood ratio tests. |
| Gaussian quadrature (Guass-Hermite) | Adaptive quadrature used by Stata and SAS 9.2 `PROC GLIMMIX` when `METHOD=QUAD` is included on the proc line. In R on the `lmer` function from the lme4 package, add `family = binary, nAGQ=n`), where n is the number of quadrature points desired (10-20 should be sufficient; Raudenbush et al., 2000; Capanu, Gönen, & Begg, 2013 show good performance with 7). | Integration over $Q^r$ quadrature points which depend on the number of random effects in the model (PQL uses only the 1 quadrature point). The 'adaptive' approach used in most software reduces the number of quadrature points based on mean and variance values (Lindstom & Bates, 1990). | Provides highly accurate estimates under a variety of conditions. Can be used for likelihood ratio tests. |
| Bayesian estimation and empirical Bayes estimates (EB: Dempster, Rubin, Tsutakawa, 1981; Lindley & Smith, 1972) | EB estimates obtained from HLM are used in the estimation of particular group intercept or slopes (e.g., plotting, or assumption checking). The REML and FIML estimation can be interpreted from an EB perspective. Statistical packages such as WinBugs (using Gibbs/MCMC) use more explicit EB approaches in the estimation process (e.g., Browne & Draper, 2006). | Fully Bayesian estimation uses Markov chain Monte Carlo (MCMC) e.g., Gibbs sampler (Geman & Geman, 1984; Zeger & Karim, 1991) | Bayes estimates are output in a "residual" file in HLM if requested under Basic Specifications. Can be used to assess normality assumption. MCMC and Bayesian approaches can produce results comparable to the higher-order Laplace or the adaptive quadrature approaches (Capano et al., 2014). |

**References**

Browne WJ, Draper D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, *1*,473–514.

Capanu, M., Gönen, M., & Begg, C. B. (2013). An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in medicine, 32*(26), 4550-4566.

Lindstrom, M. J., and Bates, D. M. (1988). Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. *Journal of the American Statistical Association, 83*, 1014-1022.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayes restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721-741.

McNeish, D. (2016). Estimation methods for mixed logistic models with few clusters. Multivariate Behavioral Research, 51(6), 790-804.

McNeish, D. (2019). Poisson multilevel models with small samples. Multivariate Behavioral Research, 54(3), 444-455.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. Educational Psychology Review, 28, 295-314.

Raudenbush, S.W., & Bryk, A.S., (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Raudenbush, S. W., Yang, M. L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of computational and Graphical Statistics, 9*, 141-157.

Wolfinger, R. (1993). Laplace's Approximation for Nonlinear Mixed Models. *Biometrika, 80*,791-795.

Zeger, S. L., & Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American statistical association, 86*(413), 79-86.