Combined shrinkage of fixed and random effects in linear mixed models using empirical Bayes

Matteo Amestoy[1], Mark van de Wiel[1], Wessel van Wieringen[1,2]

[1] Department of Epidemiology and Data Science, Amsterdam University Medical Center, Amsterdam, The Netherlands
[2] Department of Mathematics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
E-mail for correspondence: m.amestoy@amsterdamumc.nl

Abstract:
The ExposomeNL consortium aims to unravel the effect of individual and exposure-related variables on cardiovascular-related outcomes. Hereto longitudinal cohort data, combining individual-level medical records with exposure data, are available. Repeated measures over time and shared spatial information generate a complex correlation structure that is of substantive interest. While standard applications of linear mixed models (LMM) limit the number of random-effects variables, we use a high (medium) dimensional design matrix to fully capture this correlation structure. However, the high (medium) dimensionality of both the fixed- and random-effects compromise the stability of the associated estimators. We add a prior distribution to shrink both estimators. The prior's hyperparameters are estimated using an Empirical Bayes method, with an exact computation of the Hessian matrix in the Laplace approximation of the marginal likelihood. We compare the performance of our method to standard LMM algorithms using simulated data. We show that we capture complex correlation structures and improve the accuracy of the estimates. This methodology is then applied to Exposome data to analyse the combined effect of individual and exposure related variables, unravelling hitherto unknown patterns in the data that are overlooked by traditional methods.

Key words: Longitudinal data; High dimensionality; Covariance matrix; Regularization