Fixed and Random Effects Selection in Generalized Linear Mixed Effects Models

Danae Carreras-Garcia[1], *Ana Arribas-Gil*[2], *David Delgado-Gomez*[3]

[1]dcarrera@est-econ.uc3m.es, [2]aarribas@est-econ.uc3m.es,[3]ddelgado@est-econ.uc3m.es,
Department of Statistics, University Carlos III of Madrid

Generalized linear mixed effects models (GLMM) are widely used in the analysis of correlated or clustered data, such as longitudinal data for repeated measurements. They allow the inclusion of subject-specific parameters via random effects and population characteristics through fixed effects. In addition, GLMM are able to model non-continuous outcomes, which extends both generalized linear models and linear mixed models.

Maximum Likelihood estimation in GLMM has been widely discussed because the likelihood function involves an N-dimensional integral which usually can not be integrated out explicitly. A second issue lies in the fact that the covariance matrix must be positive definite, which leads to ill conditioned problems in estimation procedures when the matrix is close to singular. Another important question in GLMM is variable selection. That is the choice of a model that minimizes a certain criterion based on a trade-off between model fit, and model complexity. Most variable selection procedures are based on the inclusion of a penalization for the parameters in the optimization function. However, the majority of these methods only select significant fixed effects because in order to remove a random effect, an entire row and column of the matrix must be removed.

The goal of this study is to simultaneously select relevant fixed and random effects in GLMM through penalization. We adopt a Cholesky decomposition of the covariance matrix, that ensures the positive definiteness, thus leaving the estimation problem unconstrained. We approximate the N-dimensional integral using the Laplace approximation, and optimize the parameters with the Iterative Soft Thresholding Algorithm with Stochastic Coordinate Descent actualizations.

Keywords: Variable Selection, Regularization, Generalized Linear Mixed Effects Models.