Function selection in mixed effects models using L^1–penalization

Perrine Soret[1,2,3,4], *Cristian Meza*[5], *Karine Bertin*[5], *Marta Avalos*[1,2,3], *Philippe Hellard*[6]

[1]Univ. Bordeaux, ISPED, F–33000 Bordeaux, France
[2]INRIA SISTM Bordeaux – Sud–Ouest, F–33405 Talence, France
[3]INSERM, Centre INSERM U897–Epidémiologie–Biostatistique, F–33000 Bordeaux, France
[4]Vaccine Research Institute (VRI), F–94000 Créteil, France
[5]CIMFAV–Facultad de Ingeniería, Univ. de Valparaíso, Valparaíso, Chile
[6] French Swimming Federation, Research department, F–93500 Pantin, France

The penalization of likelihoods by L^1–norms has become an established and relatively standard technique for high-dimensional data when the assumed models are based on n independent and identically distributed observations. These techniques may improve prediction accuracy (since regularization leads to variance reduction) together with interpretability (since sparsity identifies a subset of variables with strong effects). Computationally, these penalties are attractive and their theoretical properties have been intensively studied during the last years.

Several authors have recently developed suggestions to analyze high-dimensional clustered or longitudinal data using L^1–penalization methods in mixed effects models. These approaches are mostly developed for variable selection purposes in linear and generalized linear mixed effects models (LME and GLMM, respectively) and also, but less extensive, in parametric nonlinear mixed effects (NLME) models.

Only a few works have considered the problem of selecting nonlinear functions using L^1–penalization methods in nonparametric mixed effects models, with additive (GAMM) or nonadditive predictors. Nonlinear functions are approximated by a linear combination of smooth functions (spline, wavelet or Fourier basis functions, for example) possibly combined with more irregular functions (spiky basis functions, for example). The resulting estimator depends only on a relatively small number of variables and/or a relatively small number of basis functions.

In this study we review recent advances in function selection in mixed effects models using L^1–penalization, focusing on computational strategies. We illustrate the interest of such approaches in the analysis of a twenty–year longitudinal study of training practices of elite athletes. We discuss practical aspects, including the determination of an appropriate collection of basis functions (according to the knowledge of the application field), the estimation of optimal regularization parameter (by different model selection criteria) and the choice of covariance structures.

Keywords: Lasso, Nonparametric mixed effects models, Sport science data.