

An *R* implementation of bootstrap procedures for mixed models

Jose A. Sanchez-Espigares¹ , Jordi Ocaña²

1. Department of Statistics and Operational Research. Universitat Politecnica de Catalunya

2. Department of Statistics. Universitat de Barcelona

* Contact author: josep.a.sanchez@upc.edu

Keywords: Mixed Models, Bootstrap, inference

The implementation of mixed models estimation in the *lme4* package provides a general common framework for linear, generalized and nonlinear mixed models with nested and/or (partially) crossed random effects. This package reimplements the estimation procedures for mixed models from the standard *nlme* package, Pinheiro and Bates (2000), in a more efficient way. Inference about fixed and variance components parameters can be done by means of MCMC techniques (using the *mcmc*samp method).

We present an extension of the package to include methods for generating data according an specified model and fitting it to obtain bootstrap samples of the estimators. Several bootstrap methods can be applied to generate the data: specifying the distribution for the variance components (parametric bootstrap), resampling with replacement any transformation of the random effects/residuals from the fitting process (semi-parametric bootstrap) or using extensions of the empirical distribution (wild bootstrap). Generation of data is performed keeping the design matrices for the fixed and random part of the model. First, the random effects are obtained under one of the above strategies in order to calculate the resampling linear predictor and the corresponding mean for each observation. Next, the resampling response variable is generated according to the conditional distribution considered for the response given the random effects. Trying to keep the general framework for the three kind of models, a bootstrap method based on resampling the quantile residuals, Dunn and Smyth (1996), is proposed to obtain the resampling data in the second step.

Due to the fact that estimation of generalized and nonlinear mixed models are computer intensive procedures, efficient strategies are needed to reduce the computational cost of these bootstrap methods. This work presents the features of the routines implemented and evaluates several options in comparison of Bayesian *lme4* approach and other *R* packages alternatives.

References

- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer-Verlag.
- Dunn, K. P., and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236-244.