An Alternative Package for Estimating Multivariate Generalised Linear Mixed Models in R

Rob Crouchley[1], Damon Berridge[2], Dan Grose[1]

[1]Centre for e-Science, Lancaster University, Lancaster, LA1 4YT
[2]Centre for Applied Statistics, Lancaster University, Lancaster, LA1 4YF

Contact email: r.crouchley@lancaster.ac.uk

Abstract
There are several packages at [1] that have been specially written for estimating Generalised Linear Mixed Models in R, these include, lme4 [2] and npmlreg [3]. There are also commercial systems that have algorithms for the same class of models, see e.g. Stata [4], gllamm [5] and SAS [6]. In this presentation we compare the performance of these systems with our alternative (sabreR, to be available from [7]) on some standard small to medium sized data sets and show that our alternative is very much faster. We also present a grid enabled version of the software (SabreRgrid), this shows how easy it has become to submit grid jobs from the desktop PC and the extra speed up that can be obtained by going parallel on a High Performance Computer on the grid or otherwise. This extra speed up is particularly important for estimating complex models on large and very large data sets.

SabreR is a program for the statistical analysis of multi-process event/response sequences. These responses can take the form of binary, ordinal, count and linear recurrent events. The response sequences can also be of different types (e.g. linear (wages) and binary (trade union membership)). Such multi-process data is common in many research areas, e.g. in the analysis of work and life histories from the British Household Panel Survey or the German Socio-Economic Panel Study where researchers often want to disentangle state dependence (the effect of previous responses or related outcomes) from any omitted effects that might be present in recurrent behaviour (e.g. unemployment). Understanding of the need to disentangle these generic substantive issues dates back to the study of accident proneness (Bates and Neyman, 1952) and has been discussed in many applied areas, including consumer behaviour (Massy et al, 1980) and voting behaviour (Davies and Crouchley, 1985)

SabreR can also be used to model collections of single sequences such as may occur in medical trials, e.g. headaches and epileptic seizures (Crouchley and Davies, 1999, 2001), or in single equation descriptions of cross sectional clustered data such as the educational attainment of children in schools.

We call the class of models that can be estimated by sabreR, Multivariate Generalised Linear Mixed Models. These models have special features added to the basic models to help them disentangle state dependence from the incidental parameters (omitted or unobserved effect). The incidental parameters can be treated as random or fixed, the random effects models being estimated using normal Gaussian quadrature or Adaptive Gaussian quadrature. ‘End effects' can also be added to the models to accommodate ‘stayers’ or ‘non susceptibles’. The fixed effects algorithm we have developed uses code for large sparse matrices from the Harwell Subroutine Library, see [8].

SabreR and SabreRgrid also includes the option to undertake all of the calculations using increased accuracy. This is important because numerical underflow and overflow often occur in the estimation process for models with incidental parameters. This feature does not seem to be available is other similar software [2, 3, 4, 5, 6].

References

Bates, G.E., and Neyman, J., (1952), Contributions to the theory of accident proneness, I, An optimistic model of the correlation between light and severe accidents, II, True or false contagion, *Univ Calif, Pub Stat*, 26, 705-720.

Crouchley, R. and Davies, R.B., (1999), A comparison of population average and random effect models for the analysis of longitudinal count data with base-line information, *Journal of the Royal Statistical Society, Series A*, 162, 331-347

Crouchley, R. and Davies, R.B., (2001), A comparison of GEE and random effects models for distinguishing heterogeneity, nonstationarity and state dependence in a collection of short binary event series, *Statistical Modelling*, 1, 271-285

Davies, R.B. and Crouchley, R., (1985), The determinants of party loyalty: a disaggregate analysis of panel data from the 1974 and 1979 General Elections in England, *Political Geography Quarterly*, 4, 307-320.

Massy, W.F., Montgomery, D.B., and Morrison, D.G., (1970), *Stochastic models of buying behaviour*, MIT Press, Cambridge, Mass.

URL Links

[1] http://cran.r-project.org/

[2] http://cran.r-project.org/web/packages/lme4/index.html

[3] http://cran.r-project.org/web/packages/npmlreg/index.html

[4] http://www.stata.com/

[5] http://www.gllamm.org/

[6] http://www.sas.com/

[7] http://sabre.lancs.ac.uk/

[8] http://www.cse.scitech.ac.uk/nag/hsl/