Diagnostic Tools for Linear Mixed Models

Aarathy Babu

Diagnostics in Gaussian setup

To evaluate the fitness of a statistical model, we use residual and sensitivity analyses. Residuals are used to check linearity of effects, homoskedasticity of errors, presence of outliers and influential observations etc. Sensitivity analyses (case deletion, leverage analysis etc) are undertaken to evaluate changes when some deviation is applied to the fitted model.

Residual Analysis

3 types of residuals in LMM : - Marginal residuals : predicting marginal errors - conditional residuals : predicts conditional errors - Random effects residuals : predict random effects

these lmm residuals can also confounded (cobfounded for a specific type of error if it depends on other errors than the one supposed to predict)

Global Influence Analysis

the leverage with respect to the random effects of the conditional fitted values may be confounded by the leverage with respect to the marginal fitted values.

alternative to measure leverage of the observations and units wrt random effect components = generalized random component leverage matrix.

Case deletion analysis

- Unit-oriented measures may not be convenient to detect influential units in view of the relative position of the observations within and across subjects.

Conditional Cook distance : based on observation oriented influence measures. It is a useful measure to evaluate the influence of the j -th observation from the i-th unit on the estimates.

- non-deletion method based on studentized residual sums of squares (TRSS) plots, more efficient and flexible than Cook distance-based methods to identify outlying units or observations.

Local Influence analysis

- for evaluating the changes in the analysis resulting from 'small' perturbations on the data or on some element of the model, that is to investigate the behaviour of the likelihood displacement.
- Selection of the right perturbation scheme is not that easy as arbitary perturbing a model could lead to inapproporiate inference and may lead to difficulties in the interpretation.

Alternatives

Diagnostic tools discussed soo far suggest that proposed models donot accomodate one or more features of the data. The alternatives to cope up with this problem are shown below ,

Fine tuning of the Model Good estimates of the covariance parameters might serve the (exploratory) purpose of a diagnostic. The refinement of the associated model should be employed to prevent or at least to reduce erroneous indication of influential or outlying units or within unit observations.

In the case of random polynomial coefficient models(correspond to values of powers of the time variable), the following methods are proposed which is efficient even for moderate sample sizes like 25 units with 5 repeated measures each and for non Guassian random effects or error terms:

- simple t-tests based on the estimated coefficients of standard linear regression models fitted to each unit's data as a tool for **selecting fixed effects**
- Bonferroni-corrected reference intervals for **selecting random effects**

Plots of covariances and correlations versus time between measurements (lags) may be used as a tool for identifying possible auto-regressive covariance patterns or auto correlation plots can also be used.

Elliptically Symmetric and Skew-elliptical Linear Mixed Models Linear mixed models based on elliptically symmetric or skew-elliptical distributions have been proposed as alternatives to the standard Gaussian set-up. The class of elliptically symmetric distributions includes the Gaussian, multivariate-t, power exponential etc.

Residual and leverage analyses for elliptically symmetric LMM are still not well established however the similarity with Gaussian case suggests some exploratory tools. The index plots of the weighted marginal residuals may be used to detect outliers.

The effects of asymmetry on the appropriateness of normal theory methods are, in general, more serious than those of heavy tails, and for this reason, a considerable research effort has been directed at LMM with alternative asymmetric distributions for the random effects and error terms.A Bayesian approach to deal with skew-elliptical distributions has been adopted by various authors in this context.

Robust Linear Mixed Models Robust methods for analysing LMM are considered where the covariance structure is similar for all units and wwhere the robustifying bounded functions are equal for random effects and error terms.

Computation

Specifying a model :

- construction of individual and mean profile plots. It suggests a structure for the fixed effects. . For example, the degree of a polynomial relating (at least approximately) the mean response to the time metameter (dose, for example) may help in the specification of X. The individual profiles may also suggest the degree of the unit-specific polynomials as well as possible heteroskedasticity and/or the within-unit covariance structure, giving an idea about how Z may be defined.
- modelling strategy should include : after each new model is fitted, appropriate diagnostic tools should be employed to check whether the new proposal is more adequate than the previous one.
- compare the models via criteria like AIC, BIC or likelihood ratios, when appropriate.

- For practical applications, we must replace the corresponding covariance parameters of the model with estimates and start the diagnostic procedure with an examination of plots of the modified Lesaffre–Verbeke index versus unit indices. When the proposed structure is not adequate then respecification is required.

Examples discussed

Passive Filter Example

- Aims at the prediction of random effects and shows how an heteroskedastic model identified via an analysis of the residual plots may accommodate outliers in the conditional errors and generate better predictors of random effects.

House Prices Example

- shows how some ad hoc changes in the covariance structure may cope with an apparent long-tailedness of the underlying distributions.

In the example, the authors adopted a simpler ad hoc model, introducing different compound symmetry within-unit covariance matrices (Ri) for the towns(which showed poorly fitted covariance structure), refitted the model and generated a new plot of the modified Lesaffre-Verbeke index that suggested a poorly fitted covariance structure for some additional towns.

To assess the effect of adopting a heavier tailed distribution for both the random effects and errors, authors fitted a similar model using t-distributions with estimated degrees of freedom.

Calf weight Example

- indicates how diagnostic procedures may be useful to identify and accommodate serial correlation in the conditional errors.

Procedures used in this example :

- Fitting polynomial models to longitudinal data may lead to convergence problems induced by possible ill-conditioning of the model-specification matrices (X and Z). This may be handled by choosing appropriate algorithms, by using orthogonal polynomials or by rescaling the time variable. The BFGS (Broyden–Fletcher–Goldfarb–Shanno optimisation algorithm) is used to solve the convergence issue.
- auto-correlation plot shows within-calves observations are serially correlated but does not show further serial correlation in the conditional errors.
- The corresponding plot of the modified Lesaffre–Verbeke index suggests that the covariance structure is not adequate for some animals (there might be calf-specific variability for some animals)
- d an ad hoc grouping according to the values of the modified Lesaffre–Verbeke index to avoid over-parametrisation.

Discussion

- Analysing repeated measures or longitudinal data via Gaussian LMM is convenient for various reasons. These models are very flexible (they include the class of linearisable model) , , easily interpretable and may be fitted via a series of very efficient algorithms for which software is widely available.
- if both the fixed and random components are well specified, Such models are also convenient because in addition to the population parameters, they provide insight on the covariance structure as well as on the individual components.
- Because of the flexibility, choosing appropriate models is difficult.