Zewotir's Paper

1 Demidenko's I Influence

The concept of I Influence is generalized to the non linea regression model.

1.1 Extension of techniques to LME Models

Model diagnostic techniques, well established for classical models, have since been adapted for use with linear mixed effects models. Diagnostic techniques for LME models are inevitably more difficult to implement, due to the increased complexity.

Beckman et al. (1987) applied the local influence method of Cook (1986) to the analysis of the linear mixed model.

While the concept of influence analysis is straightforward, implementation in mixed models is more complex. Update formulae for fixed effects models are available only when the covariance parameters are assumed to be known.

If the global measure suggests that the points in U are influential, the nature of that influence should be determined. In particular, the points in U can affect the following

- the estimates of fixed effects,

- the estimates of the precision of the fixed effects,

- the estimates of the covariance parameters,

- the estimates of the precision of the covariance parameters,

- fitted and predicted values.

1.2 Residual Analysis for Linear Models, LME models and GLMs

Keywords:

- Residuals (*Beginners*),

- Testing the Assumption of Normality (*Beginners*)

- Diagnostic Plots with the `plot` function

- Cook's Distance

- DFFits and DFBeta

- Standardized and Studentized Residuals

- Influence Leverage and Outlierness

1.3 Diagnostics for Random Effects

Empirical best linear unbiased predictors EBLUPS provide the a useful way of diagnosing random effects.

EBLUPs are also known as "shrinkage estimators" because they tend to be smaller than the estimated effects would be if they were computed by treating a random factor as if it was fixed (West et al)

1.4 Case Deletion Diagnostics for Mixed Models

Christensen et al. (1992) notes the case deletion diagnostics techniques have not been applied to linear mixed effects models and seeks to develop methodologies in that respect.

Christensen et al. (1992) develops these techniques in the context of REML

1.5 Methods and Measures

The key to making deletion diagnostics useable is the development of efficient computational formulas, allowing one to obtain the case deletion diagnostics by making use of basic building blocks, computed only once for the full model.

2 Tewomir - Measures of Influence in LME Models

? describes a number of approaches to model diagnostics, investigating each of the following;

- Variance components

- Fixed effects parameters

- Prediction of the response variable and of random effects

- likelihood function

? lists several established methods of analyzing influence in LME models. These methods include

- Cook's distance for LME models,

- likelihood distance,

- the variance (information) ration,

- the Cook-Weisberg statistic,

- the Andrews-Prebigon statistic.

2.0.1 Taxonomy of Cook's Distances for LMEs

? discusses a taxonomy of Cook's distance when applied to LME models.

- For variance components γ: $CD(\gamma)_i$,

- For fixed effect parameters β: $CD(\beta)_i$,

- For random effect parameters $\boldsymbol{u}$: $CD(u)_i$,

- For linear functions of $\hat{beta}$: $CD(\psi)_i$

3   Zewotir Notepad

Abstract: Linear mixed models are extremely sensitive to outlying responses and extreme points in the fixed and random effect design spaces. Few diagnostics are available in standard computing packages. We provide routine diagnostic tools, which are computationally inexpensive. The diagnostics are functions of basic building blocks: studentized residuals, error contrast matrix, and the inverse of the response variable covariance matrix. The basic building blocks are computed only once from the complete data analysis and provide information on the influence of the data on different aspects of the model fit. Numerical examples provide analysts with the complete pictures of the diagnostics.

Key words: Case deletion, influential observations, randomeffects, statistical diagnostics, variance components ratios.

Description:

The influence of observations on statistical inference is of importance in statistical data analysis. A practical and well-established approach to influence analysis is based on case deletion. We provide computationally inexpensive diagnostic tools for linear mixed models.

The diagnostics are a function of basic building blocks, computed only once from the complete data analysis, and provide information on the influence of the data on different aspects of the model fit.

Residual standard deviation.

Roy Subject effects, replicate in subject,

Cardiac data PEFR data from BLand

Royal Melbourne Hospital.

Roy demonstrates that correlation can be described under the model formulation.

$$Y_i = x.\beta + Z.u + \epsilon]$$

For the purpose of comparison of both approaches, we compute the limits of agreement for two methods described in well known data sets.

ϵ is an $n \times 1$ vector of error terms Zewotir provides routine diagnostics tools that are computationally inexpensive. $\boldsymbol{u}_i$ is a $q_i \times 1$ vector of random variables from $\mathcal{N}(0\sigma^2.I)$

4 Zewotir Measures of Influence in LME Models

4.0.2 Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

4.0.3 linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

5 Efficient Updating Theorem

Zewotir and Galpin (2005) describes the basic theorem of efficient updating.

-

$$m_i = \frac{1}{c_{ii}}$$

5.0.4 Random Effects

A large value for $CD(u)_i$ indicates that the $i-$th observation is influential in predicting random effects.

5.0.5 linear functions

$CD(\psi)_i$ does not have to be calculated unless $CD(\beta)_i$ is large.

6 Computation and Notation

with $\boldsymbol{V}$ unknown, a standard practice for estimating $\boldsymbol{X\beta}$ is the estime the variance components σ_j^2, compute an estimate for $\boldsymbol{V}$ and then compute the projector matrix A, $\boldsymbol{X\hat{\beta}} = \boldsymbol{AY}$.

? remarks that $\boldsymbol{D}$ is a block diagonal with the $i-$th block being $u\boldsymbol{I}$

7 Section 3

$$X = \begin{bmatrix} x' \\ X \end{bmatrix}$$

$$Z = \begin{bmatrix} z' \\ Z \end{bmatrix}$$

$A_{(i)}$ denote an $n \times m$ matrix A with the $i-$th row removed.

CPJ used certain statistics as the basic building blocks of case deletion diagnostics.

Theorem 2 : Basic Theorem of efficient updating (Zewotir)

8 Measures of Influence

Cook's Distance

$$CD_i(\gamma) = g'_{(i)}(I + var(\hat{\gamma})G + g$$

Large values of CD highlights observations for special attention.

Information Ratio

$$IR(\gamma) = det(I_r + var(\hat{\gamma})G$$

- $det(A)$ denotes the determinant of the square matrix A.

-

4.2. Influence on fixed effects parameter estimates

4.2.1 Analogue of Cooks Distance

The Cooks distance can be extended to measure influence on the fixed effects in the mixed models. Large values of $CD_i(\beta)$ indicates points for further consideration

4.2.2. Analogue of the variance ratio

The variance ratio measures the change of the determinant of the variance of the fixed effects parameter estimates when the i-th case is deleted.

4.2.3 Analogue of the Cook-Weisberg statistic

This statistic is used to measure the change of the confidence ellipsoid value of β.

$$\boldsymbol{y} \sim N(boldsymbol X\beta, \sigma_e^2 psilonboldsymbol H)$$

The $100(1-\alpha)\%$ confidence ellipsoid for β is....

Cook and Weisberg proposed the logarithm of the ratio $E_{(i)}$ to E as a measure of influence.

4.2.4 Analogue of the Andrews Pregibons statistic

This is another measure based on the volume of the confidence ellipsoid. AP_i

4.3 Influence on random effects prediction.

Analogue of Cooks Distance A large $CD_i(u)$ indicates that the i-th observation is influential in predicting random effects.

4.4 Influence on the likelihood function

Likelihood Distance (LD_i)

4.5 influence on the linear functions of the fixed effect parameters.

All the diagnostics are a function of the following basic building blocks 1) Studentized residuals 2) Error contrast matrix 3) The inverse of the response variable covariance matrix. The basic building blocks are computed once from the complete data set. Zewotir assumes that D is block diagonal with the i-th block being $\gamma.I$. Applications in other notepad Studentized Residuals

$$e_i^s = \frac{e_i}{s_{(i)}^2(1-h_i)}$$

where e_i^s - studentized residual $s_{(i)}^2$ - standard deviation where ith obs is deleted h_i - leverage statistic Belsley et al (1980) recommend the use of studentized Residuals to determine whether there is an outlier. Data set 1: Beckmans' aerosol data (Zewotir) high efficiency particulate air filter. toxic dust, radionuclides, and mists/ Both Beckman and CPJ rank the 14th observation as the most influential.

Data set 2: Metal oxide analysis data (Zewotir) Process and measurement variation on the properties of lots of metal oxides. type 1 and 2 chemist 1 and 2 Sample 1 and 2 Maximum likelihood variance component ratios. Full data and with cases removed.

A large number of statistics have been proposed to study the influence of individual observations in the linear mixed model. An extensive Monte Carlo simulation study is used to evaluate the appropriateness of these influence diagnostic measures. The sensitivity of the diagnostic measures to outliers and leverages is examined, and helpful results are obtained. (Zewotir and Galpin, 2006)

We consider a generalized leverage matrix useful for the identification of influential units and observations in linear mixed models and show how a decomposition of this matrix may be employed to identify high leverage points for both the marginal fitted values and the random effect component of the conditional fitted values. We illustrate the different uses of the two components of the decomposition with a simulated example as well as with a real data set.Nobre and Singer (2011)

In this paper, we consider subset deletion diagnostics for fixed effects (coefficient functions), random effects and one variance component in varying coefficient mixed models (VCMMs). Some simple updated formulas are obtained, and based on which, Cook's distance, joint influence and conditional influence are also investigated. Besides, since mean shift outlier models (MSOMs) are also efficient to detect outliers, we establish an equivalence .... (Li et al., 2009)

Margina Means (population averaged) $E[Y_{ij}]$
Conditional means (Cluster specific) $E[Y_{ij}|\gamma_i]$

References

Beckman, R., C. Nachtsheim, and R. Cook (1987). Diagnostics for mixed-model analysis of variance. *Technometrics 29*(4), 413–426.

Christensen, R., L. M. Pearson, and W. Johnson (1992). Case-deletion diagnostics for mixed models. *Technometrics 34*(1), 38–45.

Li, Z., W. Xu, and L. Zhu (2009). Influence diagnostics and outlier tests for varying coefficient mixed models. *Journal of multivariate analysis 100*(9), 2002–2017.

Nobre, J. S. and J. M. Singer (2011). Leverage analysis for linear mixed models. *Journal of Applied Statistics 38*(5), 1063–1072.

Zewotir, T. and J. S. Galpin (2005). Influence diagnostics for linear mixed models. *Journal of Data Science 3*(2), 153–177.

Zewotir, T. and J. S. Galpin (2006). Evaluation of linear mixed model case deletion diagnostic tools by monte carlo simulation. *Communications in Statistics?Simulation and Computation*® *35*(3), 645–682.