BIOS6643 Longitudinal

L9 Softwares

EJC

Department of Biostatistics & Informatics

Softwares

More detail regarding computational methods for LMM

Convergence issues, warnings and unusual estimates in SAS, PROC MIXED

Topics for this lecture:

- A Comparison of SAS versus R for fitting LMMs
- Computation methods for LMMs
- Convergence issues, warnings and unusual estimates in SAS, PROC MIXED

- **Associated reading: LMM: software and computational issues chapter**

A Comparison of SAS versus R for fitting LMMs

There are two common packages with functions that fixed mixed models: lme4 and nlme. The lme4 package has a function called lmer (stands for linear mixed-effect regression model). This function will handle many different types of random effects but does not allow for modeling of non-simple error covariance structures. However, you can fit generalized linear mixed models using the glmer function. The nlme package has the lme function that allows for modeling of both $\boldsymbol{G}$ and $\boldsymbol{R}$ matrices , although it cannot handle some more complex models very easily.

In this section we first look at a crossed random effect model using the lmer function from lme4, and then consider different covariance modeling approaches using the lme function.

Rater and subject data and the lmer function

These data were first presented in the LMM intro notes, where 4 judges (or raters) each rated 6 subjects. In one model we used subject and rater as crossed random effects. Here was the model (called 'Approach 1' in previous notes.)

$Y_{ij} = \mu + b_{iS} + b_{jR} + \epsilon_{ij}$, where i denotes subject and j denotes judge;

$b_{iS} \sim \mathcal{N}(0,\ \sigma_S^2),\ b_{jR} \sim \mathcal{N}(0,\ \sigma_R^2),\ \epsilon_{ij} \sim \mathcal{N}(0,\ \sigma_\epsilon^2)$, all independent.

Below is the SAS approach on the left, with the equivalent R approach on the right.

SAS code and output:

```
data rater; input subject rater y @@; datalines;
1 1 7 1 2 8 1 3 3 1 4 5 2 1 2 2 2 4 2 3 4 2 4 1 3 1 1 3 2
2 3 3 6 3 4 1 4 1 5 4 2 5 4 3 7 4 4 2 5 1 8 5 2 9 5 3 5 5
4 6 6 1 9 6 2 10 6 3 6 6 4 7;
proc mixed data=rater;
 class subject rater;
 model y=;
 random subject rater;
 ods output covparms=cov1; run;

Cov Parm    Estimate
subject         4.1444
rater       0.6611
Residual        3.2972

Fit Statistics
-2 Res Log Likelihood           107.2
AIC (smaller is better)     113.2

Solution for Fixed Effects

Effect  Est.   SE   DF tVal. Pr>|t|
Interc. 5.125 0.997  3  5.14  0.014
```

R code and output:

```
library(lme4)
subject=c(1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4,
5,5,5,5,6,6,6,6)
rater=c(1,2,3,4,1,2,3,4,1,2,3,4,1,2,3,4,
1,2,3,4,1,2,3,4)
y=c(7,8,3,5,2,4,4,1,1,2,6,1,5,5,7,2,
8,9,5,6,9,10,6,7)
outer=lmer(y~(1|subject)+(1|rater))
> outer
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ (1 | subject) + (1 | rater)

REML criterion at convergence: 107.2415
Random effects:
 Groups   Name        Std.Dev.
 subject  (Intercept) 2.0358
 rater    (Intercept) 0.8131
 Residual             1.8158
Number of obs: 24,
groups:  subject, 6; rater, 4
Fixed Effects:
(Intercept)
      5.125
```

Dental data and the lme function

Examples employ the lme function within the nlme package.

Data set: sample data from R, Orthodont (included with package nlme). Four variables: DISTANCE, AGE, SUBJECT, SEX. There are 4 measures on 27 subjects, at ages 8, 10, 12 and 14. The primary outcome is DISTANCE. The data is in 'data.frame' form.
Estimation method used here: REML.

Computational methods:

SAS generally uses Newton-Raphson Ridge regression. R states "The computational methods follow on the general framework of Lindstrom and Bates (1988), JASA, Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data."

Degrees of freedom:

The method for selecting denominator degrees of freedom in SAS depends on whether a RANDOM or REPEATED (or both) are included.

- For the given data and code, if there is a RANDOM statement, the 'containment' method is used (whether or not a REPEATED statement is used.
- If there is a REPEATED but no RANDOM statement, then the 'between-within' method is used.
- The DDFM option in the MODEL statement can be used to specify the DDF method, there are about 5 to choose from.

There is no mention in R about DDF

For the fixed effects other than intercept, the DDF appears to be like that of the 'between-within' method for the LME function.

The intercept DDF is different than that of any method in SAS.

For the GLS function, R appears to use the 'residual' method for DDF (since you get the same p-values in SAS when you specify DDFM=residual for Model II, and the Residual DDF is mentioned at the end of the R output).

Three models fit:

- random intercept only
- AR(1) structure only
- random intercept plus AR(1).

For models using random terms, the lme function can be used; for those without random terms but a specified R matrix (such as AR(1)), the gls function (generalized least squares) will fit the model.

Model I

Model I

SAS code and output:

```
*Model I - random intercept only;
proc mixed data=ortho;
class sex subject;
model distance = age sex / solution;
random intercept / subject=subject;
run;
```

```
The Mixed Procedure

Model Information

Covariance Structure        Variance Components
Subject Effect              Subject
Estimation Method           REML
Residual Variance Method    Profile
Fixed Effects SE Method     Model-Based
Degrees of Freedom Method   Containment

Dimensions

Covariance Parameters      2
Columns in X               4
Columns in Z Per Subject   1
Subjects                  27
Max Obs Per Subject        4
No. of Obs               108


Covariance Parameter Estimates

Cov Parm    Subject    Estimate
Intercept   Subject    3.2668
Residual               2.0495

Fit Statistics

-2 Res Log Likelihood    437.5
AIC                      441.5
AICC                     441.6
BIC                      444.1

Solution for Fixed Effects

                      Standard
Effect  Sex  Estimate  Error  DF t-Value  Pr>|t|
Intercept    17.7067  0.8339  25   21.23  <.0001
age           0.6602  0.06161 80   10.72  <.0001
Sex  Female  -2.3210  0.7614  80   -3.05  0.0031
Sex  Male     0        .      .     .      .

Type 3 Tests of Fixed Effects

         Num   Den
Effect   DF    DF   F Value   Pr > F
age      1     80   114.84    <.0001
Sex      1     80     9.29    0.0031
```

R code and output:

```
library(nlme)

#Model I – random intercept only
fm1 <- lme(distance ~ age + Sex, data = Orthodont,
random = ~ 1 | Subject)
summary(fm1)
```

```
Linear mixed-effects model fit by REML


Data: Orthodont

Fixed effects: distance ~ age + Sex


Number of Observations: 108
Number of Groups: 27


Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:   1.807425 1.431592

      AIC      BIC    logLik
 447.5125 460.7823 -218.7563

          Value Std.Error DF   t-value p-value
(Interc.) 17.706713 0.8339225 80 21.233044  0.0000
age        0.660185 0.0616059 80 10.716263  0.0000
SexFemale -2.321023 0.7614168 25 -3.048294  0.0054

 Correlation:
          (Intr) age
age       -0.813
SexFemale -0.372  0.000

Standardized Within-Group Residuals:
     Min       Q1       Med      Q3      Max
-3.74890 -0.55034 -0.02517 0.45342 3.65747
```

Note that SAS reports variances (intercept, residual), while R reports SDs.

The -2logLik is equivalent to that in SAS. But the AICs differ because R penalizes for beta parameters, SAS does not (with REML).

Model II

Model II

SAS code and output:

```
*Model II - AR(1) only;
proc mixed data=ortho;
class sex subject;
model distance = age sex / solution;
repeated / type=AR(1)
subject=subject;run;
```

The Mixed Procedure

Model Information

Covariance Structure	Autoregressive
Subject Effect	Subject
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Dimensions

Covariance Parameters	2	Columns in X	4
Columns in Z	0	Subjects	27
Max Obs Per Subject	4	No. of Obs	108

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
AR(1)	Subject	0.6259
Residual		5.2969

Fit Statistics

-2 Res Log Likelihood	445.4
AIC	449.4
AICC	449.6
BIC	452.0

Solution for Fixed Effects

Effect	Sex	Estimate	Standard Error	DF	t Value	Pr>|t|
Intercept		17.8787	1.0909	25	16.39	<.0001
age		0.6530	0.09064	80	7.20	<.0001
Sex	Female	-2.4187	0.6933	25	-3.49	0.0018
Sex	Male	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
age	1	80	51.89	<.0001
Sex	1	25	12.17	0.0018

R code and output:

```
#Model II – AR(1) structure only
fm2 <- gls(distance ~ age + Sex, data = Orthodont,
 correlation=corAR1(form =~1|Subject))
summary(fm2)
```

The GLS performed here is based on the REML likelihood by default; to use ML, add: Method="ML" as an argument in the gls function.

Generalized least squares fit by REML

Model: distance ~ age + Sex

Correlation Structure: AR(1)
 Formula: ~1 | Subject

Parameter estimate(s):
 Phi
0.6258671
Residual standard error: 2.301495
Degrees of freedom: 108 total; 105 residual

AIC	BIC	logLik
455.4483	468.7181	-222.7241

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	17.878709	1.0908637	16.389499	0e+00
age	0.652960	0.0906420	7.203723	0e+00
SexFemale	-2.418714	0.6933441	-3.488476	7e-04

Correlation:

	(Intr)	age
age	-0.914	
SexFemale	-0.259	0.000

Standardized residuals:

Min	Q1	Med	Q3	Max
-2.651488	-0.695926	-0.062146	0.486593	2.296669

Model III

Model III

SAS code and output:

```
*Model III - random int plus AR(1);
proc mixed data=ortho;
class sex subject;
model distance = age sex / solution;
random intercept / subject=subject;
repeated / type=AR(1) subject=subject;
run;
```

```
The Mixed Procedure

Model Information

Covariance Structures      Variance Components,
                           Autoregressive
Subject Effects            Subject, Subject
Estimation Method          REML
Residual Variance Method   Profile
Fixed Effects SE Method    Model-Based
Degrees of Freedom Method  Containment

Dimensions
Covariance Parameters    3
Columns in X             4
Columns in Z Per Subject 1
Subjects                 27
Max Obs Per Subject      4
No. of Obs               108
Covariance Parameter Estimates

Cov Parm   Subject   Estimate
Intercept  Subject   3.2010
AR(1)      Subject   0.05838
Residual             2.1153


Fit Statistics

-2 Res Log Likelihood  437.4   AIC   443.4
AICC                   443.6   BIC   447.3

Solution for Fixed Effects

                       Standard
Effect  Sex  Estimate  Error  DF  t Value Pr>|t|
Intercept     17.7214  0.8500 25   20.85 <.0001
age            0.6594  0.0634 80   10.40 <.0001
Sex  Female   -2.3275  0.7613 80   -3.06 0.0030
Sex  Male      0       .      .    .     .

Type 3 Tests of Fixed Effects

        Num   Den
Effect  DF    DF   F Value   Pr > F
age     1     80   108.17    <.0001
Sex     1     80   9.35      0.0030
```

R code and output:

```
#Model III – random int, plus AR(1) structure
fm3 <- lme(distance ~ age + Sex, data = Orthodont,
random = ~ 1 | Subject, correlation=corAR1())
summary(fm3)
```

```
Linear mixed-effects model fit by REML

Correlation Structure: AR(1)
 Formula: ~1 | Subject

Number of Observations: 108
Number of Groups: 27

Parameter estimate(s):
      Phi
0.05849318

Random effects:
 Formula: ~1 | Subject
       (Intercept) Residual
StdDev:   1.788899 1.454494

Data: Orthodont
      AIC      BIC    logLik
 449.3968 465.3206 -218.6984

Fixed effects: distance ~ age + Sex

            Value Std.Error DF   t-value p-value

(Interc.) 17.721416 0.8500194 80 20.848250  0.0000
age        0.659405 0.0634074 80 10.399499  0.0000
SexFemale -2.327485 0.7611852 25 -3.057711  0.0053

 Correlation:
          (Intr) age
age       -0.821
SexFemale -0.365  0.000

Standardized Within-Group Residuals:
     Min        Q1        Med        Q3       Max
-3.683027 -0.540915 -0.008097  0.461168  3.612579
```

Starting values for alpha parameters

For a numerical technique such as Newton-Raphson Ridge regression (which SAS uses in PROC MIXED), you need starting values for the $\boldsymbol{\alpha}$ parameters.

You can either specify these starting values using the PARMS statement in PROC MIXED, or use the default, which is to use the MIVQUE0 estimator values. MIVQUE0 is actually a method that can be specified as an estimation method in the PROC MIXED statement (PROC MIXED METHOD=MIVQUE0;). This is typically not done.

MIVQUE0 performs minimum variance quadratic unbiased estimation of the covariance parameters, which is a form of method of moments estimation, and it does not require an iterative method. However, simulations have shown that REML and ML are more accurate.

Nevertheless, since MIVQUE0 is based on algebraic forms and does not rely on numerical analysis, it may be useful for extremely large data sets.

Algorithms to perform ML, REML estimation

In fitting an LMM, we discussed how a ridge-stabilized Newton-Raphson algorithm is commonly used (e.g., in SAS) to maximize the likelihood with respect to the $\boldsymbol{\alpha}$ parameters. (Estimates of $\boldsymbol{\beta}$ can then be found in closed form.)

There are other computational methods that can be used to fit an LMM, including the expectation maximization (EM) algorithm, or Fisher's Scoring method.

The EM algorithm may be useful in fitting more complex LMMs such as **heterogeneity models** that allow for random terms that have non-normal distributions. [The non-normal distributions can be constructed using a mixture of normals (see Verbeke, 2000).]

The NR algorithm may not yield convergence for such models due to their complexity. The EM algorithm, which is particularly useful for ML estimation when missing data are involved. The "E step" is the expectation step; the "M step" is the maximization step. The basic steps of the EM algorithm are as follows.

1. Obtain starting values of the parameters, call it $\theta^{(1)}$.
2. *The E step*: Let y^0 denote the observed data and let $\theta^{(t)}$ denote the current value of the parameter vector theta (t=1 the first time through). Determine $E[L(\theta|y) \mid y^0,\ \theta^{(t)}]$
3. *The M step*: Determine $\boldsymbol{\theta}^{(t)}$ that maximizes $E[L(\theta|y) \mid y^0,\ \theta^{(t)}]$.
4. Repeat steps (ii) and (iii) until convergence.

The EM algorithm typically has a slow rate of convergence. Also, it is more likely to converge at a local maximum instead of global, making precision of estimates more uncertain.

It is for these reasons that the Newton-Raphson or Fisher Scoring algorithms are preferred. On the other hand, direct likelihood maximization techniques may have convergence problems for more complex models. In such cases, the EM can be considered.

While the NR algorithm uses the Hessian or observed information matrix (the matrix of second-order derivatives of the log-likelihood function), Fisher's Scoring method uses the expected information matrix, or expected Hessian matrix.

It is possible to start the numerical optimization using Fisher's Scoring method for a certain number of iterations, and then switch over to the NR method.

In PROC MIXED, including SCORING<=number> will tell SAS to use Fisher's Scoring Method up to the specified number, after which the NR algorithm will be used. For more detail, see Verbeke (2000) and the SAS Help Documentation.

Some other facts about Fisher's Scoring Method

- Yields equivalent results as 'Iteratively Reweighted Least Squares'.
- Often used to maximize Generalized linear model (GzLM) likelihoods, although the default in PROC GENMOD is once again the NR algorithm (see SAS Help Documentation).

For more use of NR, EM or Fisher's Scoring method to achieve numerical ML or REML estimates, see Verbeke (2000).

Introduction

Sometimes when fitting a linear mixed model with data you will have convergence issues. That is, the iterative numerical method used to maximize the likelihood or restricted likelihood fails to meet convergence criteria so that estimates cannot be obtained.

In other cases, you may get estimates or a partial set of estimates but you will get a warning that a problem occurred, such as a 'non-positive definite' matrix.

Some of the convergence problems are discussed in these notes. Here, I focus on PROC MIXED, although many of the same issues will face other software that you use to fit LMMs.

Fail-to-converge issues

SAS Help Documentation indicates that some reasons for non-convergence of the Newton-Raphson algorithm include flat or ridged likelihood surfaces, model misspecification or a violation of the normality assumption.

From my experiences, most of the non-convergence issues that I have run into are alleviated once I simplify the model a bit, and thus I generally attribute it to model specification.

If you do have extremely non-normal data, then you really should deal with that up front by either transforming the data so that it is more normally distributed (if possible), using a model suitable for the distribution, or identifying outliers that may be causing problems and run analyses without them. (However, I am not encouraging you to just drop the data altogether.

Ideally, if the points are real, then you want to perform analyses with and without the points; but if the model cannot handle the points, then some type of adjust may need to be made in order to perform analyses 'with them'. Or, at the very least, report the values that you were not able to fit.)

SAS states that "It is also possible for PROC MIXED to converge to a point that is not the global optimum of the likelihood, although this usually occurs only with the spatial covariance structures."

SAS lists several steps that can be taken in order to try to get the model to converge if at first you do not succeed. Many of these steps include specifying options in the optimization routine. For more details, see 'Convergence Problems' within the 'Computational Issues' page in the MIXED documentation.

Unusual estimates for covariance parameters

We know that variances should be non-negative, and that correlations should be between -1 and +1. The optimization routines that carry out likelihood maximization in PROC MIXED employ these constraints.

It is not that uncommon to see a variance estimate of 0. In terms of numerical quantities, the actual estimate would be 0 or even negative, but since there is a constraint that the variance must be nonnegative, the estimate is 0.

In practical terms, I take this to mean that based on the specified model, there is no detectable variance for the associated random effect. Note, however, that it is possible that the variance for the same random effect is positive (but not necessarily significant) if other parts of the model are changed. That's why it is important to interpret effects in relation to the model as a whole.

By default, covariance parameters are constrained in PROC MIXED optimization. Variances are not allowed to be negative, and correlations cannot have an absolute value that exceeds 1.

When you do obtain a covariance parameter estimate that is on the boundary, it suggests that the estimate using unconstrained optimization would be out-of-bounds.

For example, using the fitting an AR(1) structure for subjects as well as including a random intercept for the Ramus data yields an estimate of 0 for the variance associated with the random intercept. If you then include the NOBOUND option in the PROC MIXED statement (no slash between them), the variance estimate is a small negative number.

However, note that doing an unconstrained optimization and then setting the violating estimate to 0 will yield different estimates for other parameters in the model, relative to the constrained optimization.

Non-positive definite matrices

A matrix $\boldsymbol{M}$ is positive definite is for any $1 \times n$ real-valued vector $\boldsymbol{z}$, $\boldsymbol{z}\boldsymbol{M}\boldsymbol{z}^\top > 0$, and $\boldsymbol{M}$ is symmetric. By definition, covariance matrices are required to be positive definite. However, when fitting models, sometimes this requirement is not attained, which will either yield a warning, error or 'note' message.

A message that $\boldsymbol{G}$ is not positive definite often occurs when a variance parameter is estimated to be 0. If the associated random effect term is removed from the model or the model is simplified in some way, then the message is likely to go away. Although having a non-positive definite fitted $\boldsymbol{G}$ is not desirable, we should keep in mind that our ultimate goal is to have a realistic fitted $\boldsymbol{V}$ matrix.

Recent runs with **SAS (v9.3)** on some of my data have only given me a 'note' that $\boldsymbol{G}$ was not positive definite, and essentially removed this parameter from the model as it was not penalized for in the AIC. In addition, the fitted $\boldsymbol{V}$ matrices did seem reasonable. Thus, if direct interpretation or inference related to this parameter are not needed and the covariance structure is essentially done to account to properly handle the correlated data, then using the model with a '0' variance in $\boldsymbol{G}$ may be of practical use. Still, I would probably search for a decent comparable model for which all covariance parameters met model assumptions.

You may see a warning or error when the Hessian matrix (matrix of 2nd order derivatives of the log likelihood function), $\boldsymbol{R}$ matrix or $\boldsymbol{V}$ matrix is non-positive-definite. This might occur if there are problems with the data, such as accidentally having multiple records for a subject for the same time of measurement.

Direct quote from SAS Help Documentation: "An infinite likelihood during the iteration process means that the Newton-Raphson algorithm has stepped into a region where either the $\boldsymbol{R}$ or $\boldsymbol{V}$ matrix is nonpositive definite. This is usually no cause for concern as long as iterations continue. If PROC MIXED stops because of an infinite likelihood, recheck your model to make sure that no observations from the same subject are producing identical rows in $\boldsymbol{R}$ or $\boldsymbol{V}$ and that you have enough data to estimate the particular covariance structure you have selected. Any time that the final estimated likelihood is infinite, subsequent results should be interpreted with caution."

SAS also states that non-positive definite Hessian matrices can occur with surface saddlepoints or linear dependencies among the parameters.