MiX99

Solving Large Mixed Model Equations

MTT

Nonlinear models in MiX99

Kaarina Matilainen, MTT

MiX99 workshop 2014, Tuusula, Finland

3.12.2014

Content

- Introduction to nonlinear models
- Technical instructions and Example of threshold model
- Features of threshold model: fixed thresholds, EM vs. NR
- Check
- Possible development
- Technical instructions and Example of Gompertz function model

Introduction

- Nonlinear models in MiX99:
  - Threshold model with probit link function

$$Y_j = \text{observed response of animal } j \text{ and } y_j = \text{underlying liability variable } N(\mu_j, \sigma_j^2)$$

$$p_{ij} = P(Y_j = i) = \Phi\left(\frac{t_i - \mu_j}{\sigma_j}\right) - \Phi\left(\frac{t_{i-1} - \mu_j}{\sigma_j}\right), \quad \text{where } t' = \left(\mathrm{t}_1 < \cdots < t_m\right), t_0 = -\infty, t_m = \infty$$

  - Gompertz growth curve

$$\ln(\mathbf{y}) = \ln(\mathbf{Xb}_\alpha + \mathbf{Zu}_\alpha) - (\mathbf{Xb}_\beta + \mathbf{Zu}_\beta)\exp(\mathbf{Xb}_\kappa + \mathbf{Zu}_\kappa) + \mathbf{e}$$

- Extensions of the MiX99 program for these models are considered as test-versions.

Introduction

- Estimation of categorical variables is implemented by the generalized linear mixed model.
- Two methods:
  - Newton-Raphson as in Hoeschele et. Al (1995).
  - EM-algorithm as in Gilmour and Thompson (1998)
- Model
  - allows many linear traits and one categorical trait.
  - allows missing traits and unequal design matrices for traits.
- Thresholds can be estimated or set to be known.
- (Co)variance components have to be known on the underlying scale and residual variance of the categorical trait should be set to one.

Introduction

- Estimation of the growth curve models is base on linearization of non-linear Gompertz function model using second order Taylor series expansion (Wolfinger and Lin, 1997).
- Implemented in MiX99 as presented in Vuori et. Al (2006a) and Vuori et. Al (2006b).
- Model
  - All traits need to follow Gompertz functions.
  - Allows missing traits and unequal design matrices for traits.
- (Co)variance components could be known or estimated simultaneously by the covariance component estimation program which allows linear random regression models.

Introduction

- Nonlinear models are computer intensive: algorithms make iterations in two levels.
  - Threshol model does both iteration levels within one MiX99 run.
  - Gompertz model does only one outer level iteration and therefore MiX99 should be run repeatedly. This enables the possible update of new variance components for each iteration round.
- So far, there is no possibility to give the CLIM command file for these models yet.
- However, CLIM can be used to initialize MiX99 instruction file. Transfer a CLIM file with all traits introduced as linear traits to a MiX99 instruction file: mix99i -d instr.clm.
- Afterwards, change some lines in MiX99_DIR.DIR file.

Technical instructions with example of threshold model

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Sire evaluation with calving ease scored as 3 categories.
- 28 observations for four sires.
- Thresholds will be estimated simultaneously.
- NR-method is used for estimation.

MiX99 instruction file

- Run CLIM first: mix99i -d instr.clm
- Then, change MODEL line of the categorical trait in MiX99_DIR.DIR by giving Tn, where n is a number of thresholds, on its own column between first and second parameters.
- Right after the MODEL line, method used to analyse the threshold model must be defined: EM or NR

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- GF.clm

```
TITLE calving ease data set in Gianola and Foulley, 1983

DATAFILE   GF.dat
INTEGER    HY DAMAGE SEX SIRE
REAL       CALVEAS

PARFILE    parin
PEDFILE    GF.ped
PEDIGREE   SIRE am

MODEL
CALVEAS = HY DAMAGE SEX SIRE
```

- Execute CLIM: mix99i –d GF.clm
- Changes in MiX99_DIR.DIR

```
# MODEL: Subgr. Trait Weight  ... model factors ...
1 T2 1 - 1 2 3 4
nr
```

- mix99i < MiX99_DIR.DIR > mix99.log

- GF.dat

```
1 2 1 1 1
1 2 2 1 1
1 3 1 1 1
1 2 2 2 2
1 3 1 2 1
1 3 1 2 3
.
.
2 2 1 4 2
2 2 2 4 1
2 3 2 4 1
2 3 2 4 1
2 3 1 4 1
2 3 1 4 1
```

- GF.ped

```
1 0 0
2 0 0
3 0 0
4 0 0
```

- parin

```
1 1 1 .05
2 1 1 1
```

MiX99 solver option file

- STOP line changes to have five entries:
  1. an integer: the maximum number of PCG-iterations within each NR- or EM-round (default is 100 or number of equations in the MME)
  2. a real value: the convergence criteria value for BLUP solutions
  3. a character: the convergence criteria
  4. an enforcing character “f”
  5. an integer: the maximum number of NR- or EM-rounds (default is 5000).
- Analyse is set to be converged when only one PCG iteration round is needed within the NR round.

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- GF.slv

```
# RAM: RAM demand: H=high, M=medium, L=low
   H
# STOP: Max of BLUP iterations, Stopping criteria value, criteria, force, max of NR rounds
                     100           1.0e-4                    d       f        5000
# RESID: Calculate residuals? (Y/N)
   N
# VALID: N=no, P=prediction, S=sum of effects, Y=YD, D=DYD, I=IDD
   N
# HETVAR: adjust for heterogeneous variance (N, S, C)
   N
# TYPSOL: type of solution files? (N,Y,A)
   Y
```

- `mix99s < GF.slv >> mix99.log`

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Standard output

```
.
.
     23     0.1220E-04    0.4577E-05    0.5563E-04    0.2981E-04
     24     0.1325E-05    0.4885E-06    0.6364E-05    0.3223E-05
       rhs' * rhs =    408.013000558145
animal rhs' * rhs =    46.5782667876900
      ------------------------------------------------------
      0     0.2638E-02    0.1258E-02     0.000         0.000
     25     0.1058E-02    0.4840E-03    0.1116E-02    0.6070E-03
     26     0.1860E-03    0.2250E-03    0.3773E-03    0.1693E-03
     27     0.1607E-03    0.8531E-04    0.2971E-03   -0.1217E-03
     28     0.1074E-04    0.7404E-05    0.1644E-03   -0.8182E-04
     29     0.5714E-05    0.2356E-05    0.1938E-04    0.8552E-05
     30     0.7996E-06    0.3150E-06    0.4606E-05    0.2343E-05
       rhs' * rhs =    407.390673102719
animal rhs' * rhs =    46.5611554071602
      ------------------------------------------------------
      0     0.2039E-03    0.1049E-03     0.000         0.000
     31     0.7832E-04    0.3526E-04    0.9453E-04    0.5410E-04

   Convergence criterion reached after          31  iterations.
     Number of PCG rounds                        5
.
.
```

Solution files

- Solution files are equal to the linear mixed model case.
- When thresholds are estimated, these are printed in the output and in the end of the Solfix-file with factor name Threshold.

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Solfix

```
Fact. Trt      Level    N-Obs  Solution                 Factor Trait
   1   1           1       13 -0.31232                      HY CALVEAS
   1   1           2       15 -0.18693E-01                  HY CALVEAS
   2   1           2       13 -0.87066E-01              DAMAGE CALVEAS
   2   1           3       15 -0.21391                  DAMAGE CALVEAS
   3   1           1       14  0.39715E-01                 SEX CALVEAS
   3   1           2       14 -0.34975                     SEX CALVEAS
Fact. Trt      Level    N-Obs  Solution                 Factor Trait
   0   1           1        0  0.13872E-01           Threshold CALVEAS
   0   1           2        0  0.64919               Threshold CALVEAS
```

- Some results

$HY_2-HY_1 = 0.2936$

$DAMAGE_2-DAMAGE_1 = -0.1268$

$SEX_2-SEX_1 = -0.3895$

Probability that male calf out of the mating of the first bull to a heifer in herd 1 will experience a difficult birth is 0.1385

Some features

Fixed thresholds

- Optionally, additional characters “ft” can be specified to indicate fixed threshold values.
- In that case, a new line must follow, where the threshold values for categorical trait are defined.
- This line should contain as many real numbers as defined for the categorical trait in the MODEL line.

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Fixed threshold values from the analysis in Gianola and Foulley, 1983.
- Execute CLIM first: mix99i –d GF.clm
- Changes in the MiX99 instruction file:

```
# MODEL:
  1 T2 1 - 1 2 3 4
  nr ft
  0.375  1.011
```

- Run MiX99: mix99i < MiX99_DIR.DIR > mix99.log
  mix99s < GF.slv >> mix99.log

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Standard output

```
.
.
   Convergence criterion reached after           26  iterations.
     Number of PCG rounds                         5
.
.
```

- Some results

$HY_2-HY_1 = 0.2937$

$DAMAGE_2-DAMAGE_1 = -0.1268$

$SEX_2-SEX_1 = -0.3896$

Probability that male calf out of the mating of the first bull to a heifer in herd 1 will experience a difficult birth is 0.1384

EM vs. NR

- There should be insignificant differences in solutions between two algorithms.
- NR algorithm is critical to attain good solutions within the rounds. Increasing the maximum number of PCG-iterations within each round may lead to fewer NR rounds and in that way faster convergence finally.
- EM algorithm will need reasonable solutions to certain extent within each EM round, after which increase in accuracy will not improve the total convergence.

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Analysis done by the EM-algorithm.
- Standard output

```
.
.
    118    0.4402E-04    0.1843E-04    0.6019E-04   -0.5775E-04
    119    0.1186E-04    0.3932E-05    0.1662E-04    0.8020E-05
       rhs' * rhs =    283.049297332089
animal rhs' * rhs =    29.7429751120079
      ------------------------------------------------------
      0    0.2730E-03    0.2660E-03     0.000         0.000
    120    0.1343E-03    0.1206E-03    0.1015E-03   -0.4225E-04
    121    0.6327E-04    0.4836E-04    0.1797E-03    0.1316E-03
    122    0.2801E-04    0.1172E-04    0.3842E-04    0.3686E-04
    123    0.7587E-05    0.2515E-05    0.1057E-04   -0.5062E-05
       rhs' * rhs =    282.918159627040
animal rhs' * rhs =    29.7383764805911
      ------------------------------------------------------
      0    0.1750E-03    0.1704E-03     0.000         0.000
    124    0.8610E-04    0.7730E-04    0.6502E-04    0.2707E-04

   Convergence criterion reached after         124  iterations.
     Number of PCG rounds                       20
.
.
```

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Solfix

```
Fact. Trt      Level    N-Obs  Solution                Factor Trait
  1    1         1        13 -0.24416                      HY CALVEAS
  1    1         2        15  0.49491E-01                  HY CALVEAS
  2    1         2        13 -0.18892E-01              DAMAGE CALVEAS
  2    1         3        15 -0.14574                  DAMAGE CALVEAS
  3    1         1        14  0.10791                     SEX CALVEAS
  3    1         2        14 -0.28160                     SEX CALVEAS
Fact. Trt      Level    N-Obs  Solution                Factor Trait
  0    1         1         0  0.21840               Threshold CALVEAS
  0    1         2         0  0.85372               Threshold CALVEAS
```

- Some results

$HY_2-HY_1 = 0.2936$

$DAMAGE_2-DAMAGE_1 = -0.1268$

$SEX_2-SEX_1 = -0.3895$

Probability that male calf out of the mating of the first bull to a heifer in herd 1 will experience a difficult birth is 0.1385

Example

Simple run for Threshold model (Gianola and Foulley, 1983)

- Decrease the maximum number of PCG iterations with analysis by the EM-algorithm.
  - Change in the solver option file GF.slv:

```
# STOP: Max of BLUP iterations, Stopping criteria value, criteria, force, max of NR rounds
                         5            1.0e-4                    d       f        5000
```

- Standard output

```
 .
 .
    Convergence criterion reached after          93  iterations.
      Number of PCG rounds                       20
 .
 .
```

- Some results

HY_2-HY_1 = 0.2936

$DAMAGE_2-DAMAGE_1$ = -0.1268

SEX_2-SEX_1 = -0.3895

Probability that male calf out of the mating of the first bull to a heifer in herd 1 will experience a difficult birth is 0.1385

At the end

Check

- Model is correct.
- In case of threshold model, the possibility to reduce computing time by
  - Increasing the maximum number of PCG iterations within NR round.
  - Decreasing the maximum number of PCG iterations within EM round.

Possible development

- Introduction of these models within CLIM.
- More models allowed, e.g., several categorical traits in the model.

References

- Hoeschele, I., Tier, B., Graser, H.U., 1995. Multiple-trait genetic evaluation for one polychotomous trait and several continuous traits with missing data and unequal models. J. Anim.Sci. 73, 1609-1627.
- Gilmour, A.R., Thompson, R., 1998. Reformulated generalised linear (mixed) model aids multiple trait genetic evaluation with polychotomous calving ease. Proc. 6th WCGALP, Australia, 20, 613-616.
- Wolfinger, R.D., Lin, X., 1997. Two Taylor-series approximation methods for nonlinear mixed models. Comput. Stat. Data Anal. 25, 465-490.
- Vuori, K., Strandén, I., Mäntysaari, E.A., 2006a. Solving large scale nonlinear mixed models. In: 8th WCGALP, Brasil.
- Vuori, K., Strandén, I., Sevón-Aimonen, M.-L., Mäntysaari, E.A., 2006b. Estimation of non-linear growth models by linearization: a simulation study using a Gompertz function. Genet. Sel. Evol. 38, 343-358.
- Gianola, D., Foulley, J.L., 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15, 201-224.

Technical instructions with example of Gompertz function model

Example

Simple run for Gompertz function model (Vuori et. Al, 2006b)

- Data consisted of 15 live weight observations for each 24 animals.
- Run CLIM first: mix99i -d instr.clm
- Then, change three lines in MiX99_DIR.DIR:
  1. In the MODEL line, give G on its own column between first and second parameters.
  2. In the REGRESS line, add the table column which contains the time at measurement (or scaled time) to the end of the line.
  3. In the CVRNUM line, change the number of covariables in the covariable table to be at least the column number specified above.

Example

Simple run for Gompertz function model (Vuori et. Al, 2006b)

- G.clm

```
TITLE Growth curve data

DATAFILE  Gompertz.dat
INTEGER   GRB SIRE DAM ANIMAL DAY
REAL      LOGWEIGHT

TABLEFILE Gompertz.cov
TABLEINDEX DAY

PARFILE   parin
PEDFILE   Gompertz.ped
PEDIGREE  G am

PRECON    d  d

MODEL  RESTARTSOL
LOGWEIGHT = fixed_curve(t1 t2 t3 | GRB) G(t1 t2 t3 | ANIMAL)
```

- Gompertz.dat

```
.
.
1  1  3  6  176   5.118366
1  1  3  6  190   5.179492
1  1  3  6  204   5.241189
1  1  3  6  218   5.263865
1  1  3  6  232   5.282073
1  1  3  6  246   5.312339
1  1  4  7   50   3.384235
1  1  4  7   64   3.786631
1  1  4  7   78   4.150081
1  1  4  7   92   4.402044
1  1  4  7  106   4.599402
1  1  4  7  120   4.798263
1  1  4  7  134   4.924832
1  1  4  7  148   5.037791
.
.
```

- Gompertz.cov

```
.
.
57 1 1 1 57 0.969
58 1 1 1 58 0.986
59 1 1 1 59 1.003
60 1 1 1 60 1.02
.
.
```

- parin

```
1  1  1  100.0
1  2  1 -0.6
1  2  2 0.01
1  3  1 -0.25
1  3  2 0.00625
1  3  3 0.015625
2  1  1 0.0001
```

Example

Simple run for Gompertz function model (Vuori et. Al, 2006b)

- Execute CLIM: mix99i –d G.clm
- Changes in MiX99_DIR.DIR

```
# MODEL:
1 G 1 - 1 2 3 4
.
 .
 # REGRESS
 6   t1   t2   t3   t1   t2   t3  t5
 .
 .
 # CVRNUM
           5
```

- mix99i < MiX99_DIR.DIR > mix99.log

MiX99 solver option file

- Stopping criteria file is defined as for linear traits.
- More strict convergence criteria for BLUPs may be needed, e.g. CR < 1.0E-6.
- To decide the final convergence of the iterative process, user must define the convergence by itself over repeated BLUP analysis. One option would be to append the solutions after each round to another file which is studied for converge of solutions.

Example

Simple run for Gompertz function model (Vuori et. Al, 2006b)

- Example about script for repeated MiX99 analyses:

```
rm STOPiteration
iti=1
while [ ! -f STOPiteration -a $iti -lt 51 ] ; do
  echo "================== mix99  ===================="
  mix99i < MiX99_DIR.DIR > mix99.log$iti
  mix99s < stopping.crit >> mix99.log$iti
  cp Solunf Solold
  echo "================== checkCC ==================="
  octave -q checkCC.m
  iti=`expr $iti + 1`
done
rm Solold
```

- Eight MiX99 runs were made to obtain relative change between BLUPs less than 0.001.