Quantitative genetic (animal) model example in \( \mathbf{R} \)

Gregor Gorjanc

gregor.gorjanc@bfro.uni-lj.si

October 22, 2008

Introduction

The following is just a quick introduction to quantitative genetic model, which is usually called animal model in animal breeding scenario. This model provides inferences on parameters such as genetic (additive/breeding, dominance, ...) values and possibly also co-variance components (additive genetic variance, heritability, ...). Very nice introduction to this topic is in Mrode (2005), which also gives a list of key references. We use example from this book and will therefore be very brief.

This note is mainly for educational purposes. There are quite some programs (e.g. Druet and Ducrocq (2006) mentions \textit{ASReml}, \textit{BGF90}, \textit{DFREML}, \textit{DMU}, \textit{MATVEC}, \textit{PEST/VCE} and \textit{WOMBAT}) that can fit animal models in a general manner and we suggest to take a look at them instead of trying to reinvent the wheel in \( \mathbf{R} \).

In short animal model is an example of a mixed model:

\[
y = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},
\]

where \( y \) represents a vector of observed (measured) phenotype values, \( \mathbf{b} \) and \( \mathbf{u} \) are vectors of unknown parameters for “fixed” and “random” effects, while \( \mathbf{X} \) and \( \mathbf{Z} \) are corresponding design matrices and finally \( \mathbf{e} \) is a vector of residuals. Assuming normal density for \( y \) the following standard assumptions are taken:

\[
\begin{pmatrix}
\mathbf{y} \\
\mathbf{u} \\
\mathbf{e}
\end{pmatrix}
\sim
\mathcal{N}
\left(
\begin{pmatrix}
\mathbf{Xb} \\
0 \\
0
\end{pmatrix},
\begin{pmatrix}
\mathbf{V} & \mathbf{GZ} & \mathbf{R} \\
\mathbf{GZ}' & \mathbf{G} & 0 \\
0 & \mathbf{R} & \mathbf{R}
\end{pmatrix}
\right),
\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}
\]

Up to now all this is as in usual mixed model. Genetic aspect comes from specification of covariance matrix between elements of \( \mathbf{u} \), which usually represents sum of additive effects of genes of individuals in the pedigree. For a univariate model the covariance matrix of additive effect can be written as \( \mathbf{G} = \mathbf{A}\sigma_s^2 \), where \( \mathbf{A} \) is additive/numerator relationship matrix (Wright, 1922) and \( \sigma_s^2 \) is additive genetic variance (Falconer and Mackay, 1996).
Mixed model equations (MME)

Solution for $\mathbf{b}$ i.e. (E)BLUE and $\mathbf{u}$ i.e. (E)BLUP can be obtained from (Henderson 1949; Goldberger 1962; Henderson 1963):

$$\hat{\mathbf{b}} = (X^TV^{-1}X)^{-1} XV^{-1} \mathbf{y},$$

and

$$\hat{\mathbf{u}} = GZ^1 \left( \mathbf{y} - XV\hat{\mathbf{b}} \right),$$

but in a case with a lot of records the size of $V$ is huge and its direct inverse prohibitive if possible at all. Henderson (1950) presented the solution to this problem with so called mixed model equations:

$$\begin{pmatrix} X' R^{-1} X & X' R^{-1} Z \\ Z' R^{-1} X & Z' R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} X' R^{-1} \mathbf{y} \\ Z' R^{-1} \mathbf{y} \end{pmatrix}.$$  

Data

We will use pedigree and data example from Mrode (2005). Example shows a beef breeding scenario with 8 individuals (animals), where 5 of them have phenotype records (pre-weaning gain in kg) and 3 three of them are without records and link others through the pedigree.

```R
> library(GeneticsPed)

NOTE: THIS PACKAGE IS NOW OBSOLETE.

The R-Genetics project has developed an set of enhanced genetics packages to replace 'genetics'. Please visit the project homepage at http://rgenetics.org for information.

> data(Mrode3.1)
> (x <- Pedigree(x=Mrode3.1, subject="calf", ascendant=c("sire", "dam"),
+       ascendantSex=c("Male", "Female"), sex="sex"))

calf   sex  sire  dam  pwg
1 S4 Male S1 <NA> 4.5
2 S5 Female S3 S2 2.9
3 S6 Female S1 S2 3.9
4 S7 Male S4 S5 3.5
5 S8 Male S3 S6 5.0
The model

For this baby BLUP example we will postulate the following model:

\[ y_{ij} = s_i + a_j + e_{ij}, \]

where \( y_{ij} \) is pre-weaning gain (kg) of calf \( j \) of sex \( j \); \( s_i \) are parameters of sex effect, while \( a_j \) are parameters of additive genetic effect for pre-weaning gain and finally \( e_{ij} \) is residual. Variances for \( a_j \) and \( e_{ij} \) are assumed as \( G = A\sigma_a^2 \) with \( \sigma_a^2 = 20 \text{ kg}^2 \) and \( R = I\sigma_e^2 \) with \( \sigma_e^2 = 40 \text{ kg}^2 \).

Setting up the MME

Observed/measured phenotype records:

\[
\begin{align*}
> (y <- x$pwg)
\end{align*}
\]

\[
\begin{bmatrix}
1 & 4.5 \\
2 & 2.9 \\
3 & 3.9 \\
4 & 3.5 \\
5 & 5.0
\end{bmatrix}
\]

Design matrix (\( X \)) for sex effect:

\[
\begin{align*}
> X <- model.matrix(~ x$sex - 1)

> t(X)
\end{align*}
\]

\[
\begin{bmatrix}
1 & 2 & 3 & 4 & 5 \\
x$sexFemale & 0 & 1 & 1 & 0 & 0 \\
x$sexMale & 1 & 0 & 0 & 1 & 1
\end{bmatrix}
\]

Design matrix (\( Z \)) for additive genetic effect. Note that first three columns do not have indicators since these columns are for individuals without phenotype records and appear in the model only through the pedigree.

\[
\begin{align*}
> (Z <- model.matrix(object=x, y=x$pwg, id=x$calf))
\end{align*}
\]

\[
\begin{bmatrix}
S2 & S1 & S3 & S4 & S5 & S6 & S7 & S8 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
4 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
5 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]
Left hand side (LHS) of MME without $G^{-1}$:

```r
> LHS <- rbind(cbind(t(X) %*% X, t(X) %*% Z),
+ cbind(t(Z) %*% X, t(Z) %*% Z))
> ## or more efficiently
> (LHS <- rbind(cbind(crossprod(X), crossprod(X, Z)),
+ cbind(crossprod(Z, X), crossprod(Z))))
```

```
x$sexFemale   x$sexMale   S2  S1  S3  S4  S5  S6  S7  S8
x$sexFemale   2   0   0   0   0 1  1  0  0
x$sexMale     0   3   0   0  1  0  1  1
S2            0   0   0   0   0 0  0  0
S1            0   0   0   0   0 0  0  0
S3            0   0   0   0   0 0  0  0
S4            0   1   0   0   0 1  0  0  0
S5            1   0   0   0   0 1  0  0  0
S6            1   0   0   0   0 0  0  1  0
S7            0   1   0   0   0 0  0  1  0
S8            0   1   0   0   0 0  0  0  1
```

and adding $G^{-1}$, which is in this case $A^{-1}\alpha$ and $\alpha = \frac{\sigma^2_e}{\sigma^2_a} = \frac{40}{20} = 2$.

```r
> ## We want $A^{-1}$ for all individuals in the pedigree not only individuals
> ## with records
> x <- extend(x)
> Ainv <- inverseAdditive(x=x)
> sigma2a <- 20
> sigma2e <- 40
> alpha <- sigma2e / sigma2a
> q <- nIndividual(x)
> p <- nrow(LHS) - q
> (LHS[(p + 1):(p + q), (p + 1):(p + q)] <-
+ (LHS[(p + 1):(p + q), (p + 1):(p + q)] + Ainv * alpha)
```

```
S2  S1  S3  S4  S5  S6  S7  S8
S2  4 1.000000 1 0.000000 -2 -2 0 0
S1  1 3.666667 0 -1.333333 0 -2 0 0
S3  1 0.000000 4 0.000000 -2 1 0 -2
S4  0 -1.333333 0 4.666667 1 0 -2 0
S5 -2 0.000000 -2 1.000000 6 0 -2 0
S6 -2 -2.000000 1 0.000000 0 6 0 -2
S7  0 0.000000 0 -2.000000 -2 0 5 0
S8  0 0.000000 -2 0.000000 0 -2 0 5
```

Right hand side (RHS) of MME:
> RHS <- rbind(t(X) %*% y,
+ t(Z) %*% y)
> ## or more efficiently
> RHS <- rbind(crossprod(X, y),
+ crossprod(Z, y))
> t(RHS)

x$sexFemale x$sexMale S2 S1 S3 S4 S5 S6 S7 S8
[1,] 6.8 13 0 0 0 4.5 2.9 3.9 3.5 5

Solution

> sol <- solve(LHS) %*% RHS
> ## or more efficiently
> sol <- solve(LHS, RHS)
> t(sol)

x$sexFemale x$sexMale S2 S1 S3 S4 S5 S6 S7 S8
[1,] 3.40443 4.358502 -0.0187701 0.09844458 -0.0410842 -0.008663123 -0.1857321
S6 S7 S8
[1,] 0.1768721 -0.2494586 0.1826147

That’s all folks! Well, all for the introduction. There are numerous issues covered in the literature. A good starting point is [Mrode (2005)] as already mentioned in the beginning.

R Session information

> toLatex(sessionInfo())

- R version 2.8.0 (2008-10-20), x86_64-unknown-linux-gnu
- Locale: LC_CTYPE=en_US;LC_NUMERIC=C;LC_TIME=en_US;LC_COLLATE=en_US;LC_MONETARY=C;LC_MESSAGES=en_US;LC_PAPER=en_US;LC_NAME=C;LC_ADDRESS=C;LC_TELEPHONE=C;LC_MEASUREMENT=en_US;LC_IDENTIFICATION=C
- Base packages: base, datasets, graphics, grDevices, methods, stats, tools, utils
- Other packages: combinat 0.0-6, gdata 2.4.2, genetics 1.3.4, GeneticsPed 1.4.0, gtools 2.5.0, MASS 7.2-44, mvtnorm 0.9-2

References


