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

Gregor Gorjanc
gregor.gorjanc@bfro.uni-lj.si

October 17, 2016

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 \textbf{R}.

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

\[ y = Xb + Zu + e, \]

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

\[
\begin{pmatrix}
  y \\
  u \\
  e
\end{pmatrix}
\sim N
\begin{pmatrix}
  Xb \\
  0
\end{pmatrix}
\begin{pmatrix}
  V & \text{else} \\
  \text{else} & R
\end{pmatrix}
\]

\[ V = ZGZ' + R \]

Up to now all this is as in usual mixed model. Genetic aspect comes from specification of covariance matrix between elements of \( u \), 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 \( G = A\sigma^2_u \), where \( A \) is additive/numerator relationship matrix (Wright, 1922) and \( \sigma^2_u \) is additive genetic variance (Falconer and Mackay, 1996).
Mixed model equations (MME)

Solution for $b$ i.e. (E)BLUE and $u$ i.e. (E)BLUP can be obtained from [Henderson 1949, Goldberger 1962, Henderson 1963]:

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

and

$$
\hat{u} = GZV^{-1}(y - X\hat{b}),
$$

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:

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

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)
> 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 $G = A \sigma^2_a$ with $\sigma^2_a = 20 \, kg^2$ and $R = I \sigma^2_e$ with $\sigma^2_e = 40 \, kg^2$. 

2
Setting up the MME

Observed/measured phenotype records:

> (y <- x$pwg)

[1]  4.5  2.9  3.9  3.5  5.0

Design matrix (X) for sex effect:

> X <- model.matrix(~ x$sex - 1)
> t(X)

  1  2  3  4  5
x$sexFemale 0 1 1 0 0
x$sexMale    1 0 0 1 1
attr(,”assign”)
  [1] 1 1
attr(,”contrasts”)
attr(,”contrasts”)$‘x$sex’
  [1] “contr.treatment”

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.

> (Z <- model.matrix(object=x, y=x$pwg, id=x$calf))

  S2 S1 S3 S4 S5 S6 S7 S8
1  0  0  0  1  0  0  0  0
2  0  0  0  0  1  0  0  0
3  0  0  0  0  0  1  0  0
4  0  0  0  0  0  0  1  0
5  0  0  0  0  0  0  0  1

Left hand side (LHS) of MME without $G^{-1}$:

> 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))))
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
> ## 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:

```r
> 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

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

```
x$sexFemale x$sexMale S2 S1 S3 S4 S5
[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**

```r
> toLatex(sessionInfo())
```

- R version 3.3.1 (2016-06-21), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=en_US.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8, LC_COLLATE=C, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=en_US.UTF-8, LC_PAPER=en_US.UTF-8, LC_NAME=C, LC_ADDRESS=C, LC_TELEPHONE=C, LC_MEASUREMENT=en_US.UTF-8, LC_IDENTIFICATION=C
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: GeneticsPed 1.36.0, MASS 7.3-45
- Loaded via a namespace (and not attached): combinat 0.0-8, gdata 2.17.0, genetics 1.3.8.1, gtools 3.5.0, mvtnorm 1.0-5, tools 3.3.1

**References**


