Quantitative genetic (animal) model example in ***R***

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

October 30, 2025

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 ***ASReml***, ***BGF90***, ***DFREML***, ***DMU***, ***MATVEC***, ***PEST/VCE*** and ***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 ***R***.

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

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e},$$

where $\mathbf{y}$ 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 $\mathbf{y}$ the following standard assumptions are taken:

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

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 $\mathbf{G} = \mathbf{A}\sigma_u^2$, where $\mathbf{A}$ is additive/numerator relationship matrix (Wright, 1922) and σ_u^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):

$$\hat{\mathbf{b}} = \left(\mathbf{X}\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}\mathbf{V}^{-1}\mathbf{y},$$

and

$$\hat{\mathbf{u}} = \mathbf{G}\mathbf{Z}^{'}\mathbf{V}^{-1}\left(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}\right),$$

but in a case with a lot of records the size of $\mathbf{V}$ 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:

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

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.

```
> 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 $\mathbf{G} = \mathbf{A}\sigma_a^2$ with $\sigma_a^2 = 20\ kg^2$ and $\mathbf{R} = \mathbf{I}\sigma_e^2$ with $\sigma_e^2 = 40\ kg^2$.

Setting up the MME

Observed/measured phenotype records:

```
> (y <- x$pwg)

[1] 4.5 2.9 3.9 3.5 5.0
```

Design matrix ($\mathbf{X}$) 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 ($\mathbf{Z}$) for additive genetic effect. Note that first three columns do not have indicators since these columns are for individuals without phenotype records and apear 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 $\mathbf{G}^{-1}$:

```
> 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))))
```

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

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

```
> ## We want Ainv 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
[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 Under development (unstable) (2025-10-20 r88955), `x86_64-pc-linux-gnu`
- Locale: `LC_CTYPE=en_US.UTF-8`, `LC_NUMERIC=C`, `LC_TIME=en_GB`, `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`
- Time zone: `America/New_York`
- TZcode source: `system (glibc)`
- Running under: `Ubuntu 24.04.3 LTS`
- Matrix products: default
- BLAS: `/home/biocbuild/bbs-3.23-bioc/R/lib/libRblas.so`
- LAPACK: `/usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.12.0`
- Base packages: base, datasets, grDevices, graphics, methods, stats, utils
- Other packages: GeneticsPed 1.73.0, MASS 7.3-65
- Loaded via a namespace (and not attached): combinat 0.0-8, compiler 4.6.0, gdata 3.0.1, genetics 1.3.8.1.3, gtools 3.9.5, mvtnorm 1.3-3, tools 4.6.0

References

Druet, T. and Ducrocq, V. (2006). Innovations in software packages in quantitative genetics. In *8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, 2006-08-13/18*. Brazilian Society of Animal Breeding. Communication 27-10.

Falconer, D. S. and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Longman, Essex, U.K., 4th ed. edition.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.*, 57(6):369–375.

Henderson, C. R. (1949). Estimation of changes in herd environment. 32:709. (Abstract).

Henderson, C. R. (1950). Estimation of genetic parameters. *Ann. Math. Stat.*, 9:309.

Henderson, C. R. (1963). Selection index and expected genetic advance. In Hanson, W. D. and Robinson, H. F., editors, *Statistical genetics and plant breeding*, number 982, pages 141–163. National academy of sciences and national research council, Washington DC.

Mrode, R. A. (2005). *Linear models for the prediction of animal breeding values*. CAB International, Wallingford, Oxon OX10 8DE, UK, 2 edition.

Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.*, 56:330–338.