5.3   Pedigree BLUP

In real datasets, the assumption of unrelated sires is unrealistic, because the selection process favors that male offspring of a given sire will be selected as sires again. As a consequence, the random sire effects are not independent. The dependence structure between the sire effects must be considered in the analysis. The sire model can still be written as shown in equation (5.12), but the variance-covariance matrix of the random sire effects ($\mathbf{s}$) is no longer an identity matrix $\mathbf{I}$ times a common variance component σ_s^2, but it can be written as

$$var(\mathbf{s}) = \mathbf{A}_s * \sigma_s^2 \tag{5.13}$$

where $\mathbf{A}_s$ is the sire relationship matrix. In a dataset with q sires, the matrix $\mathbf{A}_s$ has dimensions $q \times q$ and it contains the proportions of sire effects that are passed from father to son. Together with the sire variance σ_s^2 this defines the variance-covariance structure of all sire effects. As an example, we can express the covariance $cov(s_i, s_k)$ of the sire effects between son i and its sire k, as

$$cov(s_i, s_k) = 1/2 * \sigma_s^2 \tag{5.14}$$

where the factor $1/2$ stems from the fact that sire k passes half of its genetic potential to its son i. Relating this single covariance back to the variance-covariance matrix A_s means that elements (i, k) and (k, i) are both $1/2$.

5.3.1   Example Dataset

An application of the sire model is shown in the dataset given in Table 5.6 which is taken from [Mrode, 2005]

Table 5.6: Pre-weaning Gain in kg for five beef animals

Animal	Sire	Sex	WWG
4	1	M	4.5
5	3	F	2.9
6	1	F	3.9
7	4	M	3.5
8	3	M	5.0

The objective is to predict breeding values for sires 1, 3 and 4 based on the above dataset. The trait pre-weaning gain (`WWG`) is taken as response and `Sex` is assumed to be the only fixed effect. The following values for the variance components are assumed: $\sigma_s^2 = 5$ and $\sigma_e^2 = 55$.

This type of model where the structure of the variance-covariance matrix of the random effect is given by a pedigree cannot be fit by the package `lme4`. An

extension of `lme4` is given in the R-package `pedigreemm`. In `pedigreemm` it is possible to specify the variance-covariance structure via a pedigree. For our example of the dataset in Table 5.6, this can be done as follows

```
library(pedigreemm)
ped_sire <- pedigree(sire = c(rep(NA,2), 1), dam = rep(NA,3),
                     label = as.character(c(1,3,4)))
lmem_sire <- pedigreemm(
  formula = WWG ~ Sex + (1 | Sire),
  data = tbl_sire_model,
  pedigree = list(Sire = ped_sire)
)
(smry_lmem_sire <- summary(lmem_sire))
```

```
## Linear mixed model fit by REML ['lmerpedigreemm']
## Formula: WWG ~ Sex + (1 | Sire)
##    Data: tbl_sire_model
##
## REML criterion at convergence: 8.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1180 -0.6708  0.2236  0.6708  0.8944
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Sire     (Intercept) 0.0000   0.0000
##  Residual             0.5556   0.7454
## Number of obs: 5, groups:  Sire, 3
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   3.4000     0.5270   6.451
## SexM          0.9333     0.6804   1.372
##
## Correlation of Fixed Effects:
##      (Intr)
## SexM -0.775
```

The output of `pedigreemm::pedigreemm()` is equivalent to the one given by `lme4::lmer()`. From this we can see that the sire variance results in an estimate of 0. This does not correspond to the value that, we specified as an assumption. This has two reasons. The assumed sire variance was not estimated from the small dataset in Table 5.6, but from a larger dataset not shown here. The second reason is that it is not possible with `pedigreemm::pedigreemm()` to use an assumed variance component as input. The breeding values for the sires can be obtained by the function call

```
ranef(lmem_sire)
```

```
## $Sire
##   (Intercept)
## 1           0
## 3           0
## 4           0
```

The predicted sire breeding values are also all equal to 0. The reason for this is that the sire variance component was estimated to be 0. Hence for small datasets, `pedigreemm::pedigreemm()` cannot be used.

5.3.2   Mixed Model Equations

In a series of papers ([Henderson, 1953], [Henderson, 1963] and [Henderson, 1975]) which are summarized in [Henderson, 1982], a technique called **mixed model equations** was developed to solve for solutions of predicted values in a linear mixed effects model. For a given linear mixed effects model with $var(e) = I * \sigma_e^2$ and $var(s) = A_s * \sigma_s^2$

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e} \tag{5.15}$$

the solutions for the fixed effects estimates and the predicted values of the random effects can be obtained by solving the following set of equations.

$$\left[\begin{array}{cc} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda A_s^{-1} \end{array} \right] \left[\begin{array}{c} \hat{b} \\ \hat{s} \end{array} \right] = \left[\begin{array}{c} X^T y \\ Z^T y \end{array} \right] \tag{5.16}$$

where $\lambda = \sigma_e^2/\sigma_s^2$. For our example the matrices X and Z are defined as

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

and

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

The matrix A_s^{-1} can be obtained by calling the function `pedigreemm::getAinv()` on the defined sire pedigree.

```
mat_Ainv <- pedigreemm::getAInv(ped = ped_sire)
```

That results in

$$A_s^{-1} = \begin{bmatrix} 1.33333333333333 & 0 & -0.666666666666667 \\ 0 & 1 & 0 \\ -0.666666666666667 & 0 & 1.33333333333333 \end{bmatrix}$$

The vector y corresponds to the vector of observations. With that we can solve the mixed model equations.

For the fixed effects we get

Table 5.7: Solutions for fixed Effect of Sex

Sex	Solution
F	3.381986
M	4.335671

For the random sire breeding values, we get

Table 5.8: Solutions for random Breeding Values of Sires

Sire	Solution
1	0.0220022
3	0.0140264
4	-0.0430418